

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 7.40034 Seconds

(without alignments)
2028.884 Million cell updates/sec

Title: US-09-697-863A-2
Perfect score: 1902
Sequence: 1 MELGSCLEGGREAAEEEGEP.....DCGRFPSDHWGLCNLDIIL 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	140	7.4	429	1	NOCT_MOUSE	O35710 mus musculu
2	134.5	7.1	388	1	NOCT_XENLA	P79942 xenopus lae
3	128	6.7	431	1	NOCT_HUMAN	Q9uk39 homo sapien
4	110.5	5.8	361	1	APEA_DICDI	P51173 dictyostell
5	108	5.7	477	1	ISCI_YEAST	P40015 saccharomyc
6	105.5	5.5	253	1	YBHP_ECOLI	P75772 escherichia
7	105.5	5.5	670	1	Y759_HUMAN	Q9unk9 homo sapien
8	103	5.4	548	1	G6PI_YERPE	Q8zas2 yersinia pe
9	102.5	5.4	267	1	EX3_HAELIN	P44318 haemophilus
10	102.5	5.4	549	1	G6PI_SALTI	O8zi17 salmonella
11	102.5	5.4	549	1	G6PI_SALTY	O8zi17 salmonella
12	102	5.4	1391	1	RPOB_MYCPN	P78013 mycoplasma
13	99.5	5.2	333	1	PHL1_BACCE	P09599 bacillus ce
14	98.5	5.2	549	1	G6PI_HAETN	P44312 haemophilus
15	98.5	5.2	3396	1	PGCV_HUMAN	P13611 homo sapien
16	98	5.2	253	1	NOCT_RAT	Q9et55 rattus norv
17	97.5	5.1	419	1	NSMA_MOUSE	O70572 mus musculu
18	97.5	5.1	549	1	G6PI_PASMU	Q9cul2 pasteurella
19	96.5	5.1	422	1	NSMA_RAT	Q9et64 rattus norv
20	96.5	5.1	996	1	OSH3_YEAST	P38713 saccharomyc
21	93.5	4.9	295	1	YCX7_EUGGR	P31920 euglena gra
22	93.5	4.9	856	1	ENV_HYIMN	P05877 human immun
23	93	4.9	549	1	G6PI_ECOLI	P11537 escherichia
24	93	4.9	1259	1	LINI_HUMAN	P08547 homo sapien
25	92.5	4.9	556	1	PHL_LEPIN	P17627 leptospira
26	92.5	4.9	565	1	NPH1_CANFA	Q9tul9 canis famil
27	91.5	4.8	807	1	MCM3_XENLA	P49739 xenopus lae
28	90.5	4.8	193	1	IL18_HUMAN	Q14116 homo sapien
29	90.5	4.8	657	1	HUTH_HUMAN	P42357 homo sapien
30	90	4.7	507	1	RRPP_MEASE	P03422 measles vir
31	90	4.7	849	1	VNCS_AEDEV	P27454 aedes denso
32	89.5	4.7	587	1	VN04_YEAST	P53933 saccharomyc
33	89.5	4.7	2663	1	CENE_HUMAN	Q02224 homo sapien

34	89	4.7	246	1	YSW4_CAEEL	Q10019 caenorhabdi
35	89	4.7	746	1	STM2_HUMAN	Q9p246 homo sapien
36	88.5	4.7	498	1	FLID_BACSU	P39738 bacillus su
37	88.5	4.7	657	1	HUTH_MOUSE	P35492 mus musculu
38	88.5	4.7	2290	1	POLG_EMCV	P03304 encephalomy
39	88	4.6	387	1	GMDS_DROME	Q9vmw9 drosophila
40	88	4.6	862	1	PGCV_MACNE	Q28858 macaca neme
41	88	4.6	1125	1	MAD4_MOUSE	P27546 mus musculu
42	88	4.6	3678	1	DMD_MOUSE	P11531 mus musculu
43	87.5	4.6	408	1	PGK_SULISO	P50317 sulfolobus
44	87.5	4.6	462	1	YMA5_YEAST	Q04235 saccharomyc
45	87.5	4.6	558	1	G6PI_CRIGR	P50309 cricetulus

ALIGNMENTS

RESULT 1
ID NOCT_MOUSE STANDARD; PRT; 429 AA.
AC O35710; Q9QZG9;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nocturnin (CCR4 protein homolog).
GN CCRN4L OR NOC OR CCR4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=99453012; PubMed=10521507;
RA Dupressoir A., Barbot W., Loireau M.P., Heidmann T.;
RT "Characterization of a mammalian gene related to the yeast CCR4
general transcription factor and revealed by transposon insertion.";
RL J. Biol. Chem. 274:31068-31075(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Retina;
RX Wang Y., Osterbur D.L., Green C.B., Besharse J.C.;
RA "Mammalian homologs of Xenopus nocturnin: conservation of structure
and circadian regulation";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 62-429 FROM N.A.
RC STRAIN=DBA/2J; TISSUE=Liver;
RX MEDLINE=97190339; PubMed=9038221;
RA Puech A., Dupressoir A., Loireau M.P., Mattei M.-G., Heidmann T.;
RT "Characterization of two age-induced intracisternal A-particle-related
transcripts in the mouse liver. Transcriptional read-through into an
open reading frame-with-similarities to the yeast ccr4 transcription
factor";
RL J. Biol. Chem. 272:5995-6003(1997).
CC -!- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR
OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH
MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT
LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CCR4/NOCTURNIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF183960; AAD56547.1; -;
DR EMBL; AF199491; AAG01384.1; -;
DR EMBL; U70139; AAB62717.1; ALT_FRAME.
DR MGD; MGI:109382; Ccr4.

	YOR -> LPA (IN REF. 3).
DR InterPro; IPR005135; Exo_endo_phos.	
DR Pfam; PF03372; Exo_endo_phos; 2.	
Biochemical rhythms.	
FT CONFLICT 123 125	
SEQUENCE 429 AA; 48300 MW; CB9FB55D84E13942 CRC64;	

Query Match	7.48;	Score 140;	DB 1;	Length 429;
Best Local Similarity	21.78;	Pred. No. 0.0006;		
Matches	89;	Conservative	55;	Mismatches 163;
				Indels 104;
				Gaps 19;

```
QY      12 EAAEEGEPEYK-----RLLCVEFASVSCDAVAQCFLAENDWEMERALNSY 61
        :| | : | | : | | : | | :
Db      47 QAAAASGARSRPRTVSSMGNGTSRLYSALAKTYNASSAAQHPEYLVDPE-----H 99
```

```

QY      62 FEB-----ESALERRPETISEPKTYVDLTNEETITSTSKISPSEDTQENGSMF 114
        | | | : | | | : ||| | | : |
DB      100 LEPIDPKELLECRAVLHTRPRYQ---RDFVDL---RTQSSSH-----SPI 141

```

QY 115 SLITWNI-----DGLD-----LNNLSERARGVCSYLALXSPDIVFLQEIVIPYYSY 160
::|||::|||::|||::|||::|||:
Db 142 RVMQWNILAQALGEGKDNFVQCPEALKWEERKCLILEETLAYQPDIICLQEV-DHYEDT 2000

```
QY 161 LKRRSSNEYITGHEEGYF-----TAINLKRSRVKLKSGEIIPEPST 202
      : | :: : | : : | | | : :
Db 201 FQPLLSR----LGYQGTFFPKPWSPCLDVEHNNGPDCALFELQNRKLIISSTNIRLTAM 256
```

```

QY 203 KMRRLLCV-----HVNVSQNELCMTLSHLESTRGHAERMQLKVLKRMQZAPESATV- 257
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 257 TLKTNQVAIAQTLCEKESGRQFCIAVTHLKARTGWERFRSAQCDLLQNLQNTQCAKIP 316

```

```

QY 258 -IFAGDTNRDRE--VTRCGGLPNNIVDWEFL--GKPKHCQYIWDIQMNSNLGLIAC 317
      |  || | |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 317 LIVCGDFNAEPTLEEYKHFASSSLNLSNAIKLLSPDGQSEPPYTTW-----KIRTSGEC 370

```

```
QY      312 KLRFDRIEAKAAEGHILPRSDLL-----GLENDGGRPSDHWGLCN   357
          : | |: | | | | | | | | | | | | | | | | | | | |
Db      371 RHTLDYIWSKHALS--VTSALDLITEEQIGPNRLPSFHYPSDHLSLVCD  418
```

RESULT 2	
NOCT_XENLA	
ID NOCT XENLA	STANDARD;
	PRT; 388 AA.

DT	15-JUL-1999	(Rel. 38, Created)
DT	15-JUL-1999	(Rel. 38, Last sequence update)
DT	15-JUN-2002	(Rel. 41, Last annotation update)

05 Xenopus laevis (African clawed frog).
0C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
0C Amphibia; Batrachia; Anura; Mesobatrachia; Pipioidea; Pipidae;

```

OX      NCBI_TaxID=8355;
RN      [1]
RP      SEQUENCE FROM N.A.

```

RX	MEDLINE=97121484; PubMed=8962150;
RA	Green C.B., Besharse J.C.;
RT	"Identification of a novel verteb

Proc. Natl. Acad. Sci. U.S.A. 93:14884-14888(1996).
-!- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR
OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH

CC LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES.
CC -I- TISSUE SPECIFICITY: EXPRESSED ONLY IN THE PHOTORECEPTORS OF THE
CC RETINA. EXPRESSION IS CONTROLLED BY THE RETINAL CIRCADIAN CLOCK

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U74761; AAB39495.1; -;
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.

Query Match 7.18; Score 134.5; DB 1; Length 388;

Matches 77; Conservative 58; Mismatches 144; Indels 109; Gaps 15;
QY 33 ASVASCDAAVACFLAENDWEMERALNSYFEPVE-----ESALLERPETISEPKTYVDL 87

Db 36 AKTLSSSAVSQELLEASQHDQSEPLD-----PKELLDECQVALQDRPARLH--RDFTSL 88
QY 88 TNEETDSTTSKISPSEDITQQENGSMFLITWNI-----DG-----LDLNLISERA 133

Db 89 RSESSSQPT-----FRVMQWNILAQALGEGKDNFIMCPMEALKWEERK 133

QY 134 RGVCSYLALYSPDVIQLQEV-----IPPYYS-----YLKRRSSNYEIIITGHEGYFT 180

Db 134 YLLEETLMYQPDVLCLOEVDHYEDTFQPILSRLGYQCTFLAKPWPSPCLDVEHNNGPDCG 193

QY 181 AIMLKSRVKLKSEQEIIIPSPSTKMNRLLCVHNV---SGNELCLMTSHLESTRGHAEE 236

Db 194 ALFFLQDRFQVLNSAKIRLSARTLKTNQVAIAETLQCCETIGRQJCEAVTHLKARTIGWERF 253

QY 237 RMNQLKMKVLMQEAPE\$ATV--IFAGDTNLRDRE--VTRCGGLPNNIVDVWEFL----- 287

Db 254 RLAQGSDDLNDLESITQGAIVPLILICGDFNADPTELEVYKRRASSSLNLSAIKLLSEDEG 313

QY 288 -----GKPKH-CQYTWDTQ----MNSNLGITAAACKLRFDRIFFRAAAEEGHI 329

DB 314 SEPPYTWIKIRIGESCHILDIWISQHALRVNAALDLE 324
 QY 330 IPRSLDLGLEKLDGCRFPDHWGLLN 357

303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1

ID	NOCT_HUMAN	STANDARD;	PRT;	431 AA.
AC	Q9UK39; Q9HD93; Q9HD94; Q9HD95;			
DT	15-JUN-2002 (Rel. 41, Created)			

CCRN4L OR NOC OR CCR4.
GN

OC *nono* *Septans* (*nono*);
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 OX

RA Dupressoir A., Barbot W., Loireau
RX MEDLINE=99453012; PubMed=10521507;
RP SEQUENCE FROM N.A.

RT	general transcription factor and revealed by transposon insertion.";
RL	J. Biol. Chem. 274:31068-31075(1999).
RN	[2]

RA Wang Y., Osterbur D.L., Green C.B., Besharse J.C.;
RT "Mammalian homologs of *Xenopus* nocturnin: conservation of structure
RT and circadian regulation.";

CC -1- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR
CC OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH
CC MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT

CC MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT
CC LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CCR4/NOCOTURIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF183961; AAD56548.1; -
DR EMBL; AF199492; AAG01387.1; -
DR EMBL; AF199493; AAG01388.1; -
DR EMBL; AF199494; AAG01389.1; -
DR Genew; HGNC:14254; CCRNAL.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 3.
KW Biological rhytms.
FT CONFLICT 69 69 T -> N (IN REF. 2).
FT CONFLICT 77 77 G -> A (IN REF. 2).
FT CONFLICT 266 266 A -> T (IN REF. 2; AAG01389).
FT CONFLICT 341 341 N -> S (IN REF. 2).
SQ SEQUENCE 431 AA; 48150 MW; B61EF484E8D29AF5 CRC64;

Query Match 6.7%; Score 128; DB 1; Length 431;
Best Local Similarity 20.4%; Pred. No. 0.0057;
Matches 86; Conservative 56; Mismatches 156; Indels 124; Gaps 20;

OY 20 PEVKKRRLLCYEFA---SVASCDAAVAQCFLAENDWE---MERALNSYFEPVEESALE 72
Db 39 PRPASPRLLAAASASGARSCSRIV--CSMGTGTSRLYSGIAKTLS-----SAAS 88
OY 73 RRPEITSEP-----KTYVDLTNEETDSTTSKISPS 103
Db 89 QHPEYLVSPDEHLEPIDPKELLECRVAVLHTRPPRFQDFVDLRTD--CPSTHPI--- 143
OY 104 EDTQENGSMFLITWNI-----DGID-----LNNLSERARGVCSYLALYSPDYIF 149
Db 144 -----RVMQWNILAQALGEGKDNFVQCPVEALKWEKCKLLEFILLAYQPDILC 192
OY 150 LQEVIPYYSYLKRRSSNYEITGHEGYF-----TAIMLKSRVKL 191
Db 193 LQEV-DHYFDTFQPLSR----LGYGTFPPKPMSPCLDVEHNNNGPDGALFFLQNRKFL 247
OY 192 KSOEIIPEPSTKMRNLLCV----HVNVSQNELCLMTSLSTRGHAAERMNQLKMYLKK 247
Db 248 VNSANIRLTAMTLKTNOVAIAQTLCKESGRQFCIAVTHLKARTGWERFRSAQGDLLQN 307
OY 248 MQAPESATV--IFAGDTNLR-DREYTRCGGLPN-NIVDWEFL--GPKHCOYTWDTO 300
Db 308 LQNTYGAKIPLIVCGDFNAEPTEYYKHPASSNLNLNSAYKLLSADGQSEPPYTTW--- 364
OY 301 MNSNLGITACKLRFDRIFRRAAEEGHIIPRSLDL-----GLEKLDGGRFSDHWGLL 355
Db 365 ---KIRTSGECHRTLDIWYSKHALN---VRSALDLTTEOIGPNRLPSFNYPDHLSTV 418
OY 356 CN 357
Db 419 CD 420

RESULT 4
APEA_DICDI
ID APEA_DICDI STANDARD; PRT; 361 AA.
AC P51173;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) (Class II
DE apurinic/apyrimidinic(AP)-endonuclease).
GN APEA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
OX NCBI_Taxid=44689;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96226184; PubMed=8657579;
RA Freehand T.M., Guyer R.B., Ling A.Z., Deering R.A.;
RT "Apurinic/apyrimidinic (AP) endonuclease from Dictyostelium
RT discoideum: cloning, nucleotide sequence and induction by sublethal
RT levels of DNA damaging agents."
RL Nucleic Acids Res. 24:1950-1953(1996).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- INDUCTION: BY DNA-DAMAGING AGENTS INCLUDING ULTRAVIOLET LIGHT,
CC MNNG, GAMMA RAYS, BLEOMYCIN, AND STREPTOZOTOCIN.
CC -1- SIMILARITY: BELONGS TO THE AP/EXO A FAMILY OF DNA REPAIR ENZYMES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Query Match 5.8%; Score 110.5; DB 1; Length 361;
Best Local Similarity 21.5%; Pred. No. 0.12;
Matches 62; Conservative 52; Mismatches 94; Indels 81; Gaps 15;

OY 8 EGGREAAEEGEPEVKKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNSYFEPVE 67
Db 27 EEEKEEVEEE-EEDDKRRLVKKTPAKA-----PAK 57
OY 68 ESALERRPETISEPKTYVDLTNETTSTTSKISPSEDTQOE---NGSMFLITWNIDGL 124
Db 58 KAAAKKSK--DEDEDEEKEEETNTKTASVSTAIDNLDPEKVEENQMKIISWVAGF 115
OY 125 DLNNLSERARGVCSYLALYSPDYIFLQEV-IPYYSYLKK---RSSNYEITGHEGYF 179
Db 116 K----SVLSKGFTECEKENPDVLCIQETKINP--SNIKKDQMPKGYEHFIEADQKGHH 169
OY 180 -TAIMLKSRVKLKSOEIIPEPSTKMRNLLCVHVNVSQNELCL-----MTSHLES-- 229
Db 170 GTGVLTKK-----PNAITFGIGIAKHDN-EGRVITLLEYQFYIVNTYIPNAG 216
OY 230 TRGHAERMNQLKMYLKK-----MQAPESATVIFAGDTNLRDREV 270
Db 217 TRG-----LQRLDYRIKENVDFQAYVLEKLNATKPIIWCGLNVAHTEI 260

RESULT 5
ISCL_YEAST

```
ID ISCL YEAST STANDARD; PRT; 477 AA.
AC P40015;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE inositol phosphogolipids phospholipase C (EC 3.1.4.-) (IPS
DE phospholipase C) (IPS-PLC) (Neutral sphingomyelinase) (N-SMase)
DE (nsmase).
GN ISCL OR YER019W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=20564359; PubMed=11006294;
RA Sawal H., Okamoto Y., Luberto C., Mao C., Bielawska A., Domae N.,
RA Hannun Y.A.;
RT "Identification of ISCL (YER019W) as inositol phosphosphingolipid
phospholipase C in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 275:39793-39798(2000).
CC -1- FUNCTION: Responsible for the hydrolysis of the
phosphosphingolipids (IPS), inositol phosphorylceramide (IPC),
mannosylinositol phosphorylceramide (MIPC), and mannosyldiinositol
phosphorylceramide (M(IP)2C). Also active on sphingomyelin, but
this activity is probably not physiologically relevant.
CC -1- COFACTOR: Magnesium.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE NEUTRAL SPHINGOMYELINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18778; AAB64552.1; -.
DR SGD; S0000821; ISCL.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos. 1.
KW Hydrolase; Magnesium; Transmembrane.
FT TRANSMEM 399 417 POTENTIAL.
FT TRANSMEM 425 449 POTENTIAL.
FT METAL 100 100 MAGNESIUM (BY SIMILARITY).
FT SITE 233 233 IMPORTANT FOR SUBSTRATE RECOGNITION (BY
SIMILARITY).
FT ACT_SITE 334 334 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 477 AA; 53940 MW; 0670FD303FEB8EFF CRC64;
```

Query Match 5.7%; Score 108; DB 1; Length 477;
Best Local Similarity 22.0%; Pred. No. 0.28;
Matches 78; Conservative 47; Mismatches 124; Indels 106; Gaps 20;

```
QY 90 EETTDSTKISPSEDTQENGSMFSL-ITWNIDGLDL--NNLSERARGVCSYLALYSP 145
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 13 KEDGQSEFEALNGTNAIMSDNSKAYSIKFLFNWTWGLKYVSKHRKERLRAIADKLGHSM 72

QY 146 -----DVIFLQEV-IPPYYSYLKKR-SSNYEILITGHEGYFT-- 180
   ||| ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 73 LTPISDELLPNGGDSNENEDYDIALQEIWCVEDWKYLLASACASKPYQRLFHSGILTGP 132
```

```
QY 181 --AIMLKRSRVKLKSOEIIFFP-----STKMRNLICVHNVSGNELCLMTSHL 227
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 133 GLAIL--SKVPIESTFLYRFPINGRPSAVFRGDWYVGKSLAITVLNTGTRPIAIMNSHM 189

QY 228 E---STRGHAERMNQ-----LKNVLKKMQAPEASATYIFAGDTNLRDREVTTCGLPN 278
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 190 HAPYAKQGDAAVYLCHRSCQAWDFSRLLIKLYROA--GYAVIYVDLN-----SRPESLPH 241

QY 279 -----NIYDVWEFL-GKP-----KHQ-----YTWDTQMSNLGITA 309
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 242 KFLTQEAGLYDSWEQLHGKODLAVIARLSPLOQLLKGTCTCDLSLNTWRAQRQPD----E 297

QY 310 ACKLRF---DRIFRAAAEEGHIIPRSLDLGLEKIDCGRPSDHWGLLCNLDIT 361
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 298 ACRLDYALIDPDPLQTVDAVRFTER-----IPLDCS--VSDHFAYSCTLNIV 344

RESULT 6
YBHP_ECOLI STANDARD; PRT; 253 AA.
ID YBHP_ECOLI
AC P75772;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ybhp.
GN YBHP OR B0790 OR Z1009 OR ECS0868.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimelanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
```



```

RA  Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to licensee@isb-sib.ch).
CC  -----
DR  EMBL; AE000181; AAC73877.1; -.
DR  EMBL; D90716; BAA35449.1; -.
DR  EMBL; AE005259; AAG55161.1; -.
DR  EMBL; AP002553; BAB34291.1; -.
DR  EcoGene; EG13672; ybhp.
DR  InterPro; IPR005135; Exo_endo_phos.
DR  Pfam; PF03372; Exo_endo_phos; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 253 AA; 28790 MW; DEB29724626A167B CRC64;

```

	Query Match	5.58;	Score 105.5;	DB 1;	length 253;
	Best Local Similarity	20.8%;	Pred. No. 0.18;		
	Matches 59;	Conservative 33;	Mismatches 82;	Indels 109;	Gaps 11;
QY	104 EDTQÖENGSMFSLITWNID-----GLDLNLTSEARGVCSYLALYSVDVIFLQEV--	153			
Db	3 DQTQÖFS--FKVLTFINHHKGFYAFNRRIPLPELRDAVRTV-----SADIVCLQÖVMG	52			
QY	154 ---IPPYSYLLKRSSNYEIIITG-----HEGGYFTAILMKSRVKLSQE--	195			
Db	53 AHEVHPLHVENPDTSHYEFLADTMWSDFAYG RNAVPDEGHGNNAV L--SRYPTEHYENR	110			
QY	196 -----IIEPSTKMRRNLCCVHNVSCHNELCLMTSHLESTRGHAERMN	239			
Db	111 DVSVDGAEEKRGVLYCRIVP-PMTGKAIHVMCVHLGL-----REAHRRQA	152			
QY	240 QLKMLVKKMQEAPESATVIFAGDTNLRDRREVTRCGGLPNNIIVDWVEFLGPKPHCQYTWDT	299			
Db	153 QLAMLAEWVNELPDGEPPVLYAGDFN-----DWRQ	181			
QY	300 QMNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLLGLEKL	342			
Db	182 KANHPLKVQAG---LDEIFTTRAHG RPARTFPVOEPPLRLDRI	220			

RESULT 7			
ID	Y759_HUMAN	STANDARD:	PRT; 670 AA.
AC	O9UNK9; O94859;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Hypothetical protein KIAA0759.		
GN	KIAA0759.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=99087487; PubMed=9872452;		
RA	Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,		
RA	Tanaka A., Kotani H., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. XI.		
RT	The complete sequences of 100 new cDNA clones from brain which code		
RT	for large proteins in vitro.";		
RL	DNA Res. 5:277-286(1998).		
RN	[2]		

```

RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rowen L.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE CCR4/NOCTURIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; AB018302; BAA34479.1; ALT_INIT.
DR EMBL; AL137268; CAB70667.1; ALT_INIT.
DR EMBL; AF111169; AAD44362.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 2.
KW Hypothetical protein.
SQ SEQUENCE 670 AA; 75275 MW; 3D40E26813E27D83 CRC64;

```

Query Match	Best Local Similarity	Score	DB 1;	Length	DB 2;	Score	DB 3;	Length	DB 4;	Score	DB 5;	Length	DB 6;	Score	DB 7;	Length	DB 8;	Score	DB 9;	Length	DB 10;	Score	DB 11;	Length	DB 12;	Score	DB 13;	Length	DB 14;	Score	DB 15;	Length	DB 16;	Score	DB 17;	Length	DB 18;	Score	DB 19;	Length	DB 20;	Score	DB 21;	Length	DB 22;	Score	DB 23;	Length	DB 24;	Score	DB 25;	Length	DB 26;	Score	DB 27;	Length	DB 28;	Score	DB 29;	Length	DB 30;	Score	DB 31;	Length	DB 32;	Score	DB 33;	Length	DB 34;	Score	DB 35;	Length	DB 36;	Score	DB 37;	Length	DB 38;	Score	DB 39;	Length	DB 40;	Score	DB 41;	Length	DB 42;	Score	DB 43;	Length	DB 44;	Score	DB 45;	Length	DB 46;	Score	DB 47;	Length	DB 48;	Score	DB 49;	Length	DB 50;	Score	DB 51;	Length	DB 52;	Score	DB 53;	Length	DB 54;	Score	DB 55;	Length	DB 56;	Score	DB 57;	Length	DB 58;	Score	DB 59;	Length	DB 60;	Score	DB 61;	Length	DB 62;	Score	DB 63;	Length	DB 64;	Score	DB 65;	Length	DB 66;	Score	DB 67;	Length	DB 68;	Score	DB 69;	Length	DB 70;	Score	DB 71;	Length	DB 72;	Score	DB 73;	Length	DB 74;	Score	DB 75;	Length	DB 76;	Score	DB 77;	Length	DB 78;	Score	DB 79;	Length	DB 80;	Score	DB 81;	Length	DB 82;	Score	DB 83;	Length	DB 84;	Score	DB 85;	Length	DB 86;	Score	DB 87;	Length	DB 88;	Score	DB 89;	Length	DB 90;	Score	DB 91;	Length	DB 92;	Score	DB 93;	Length	DB 94;	Score	DB 95;	Length	DB 96;	Score	DB 97;	Length	DB 98;	Score	DB 99;	Length	DB 100;	Score	DB 101;	Length	DB 102;	Score	DB 103;	Length	DB 104;	Score	DB 105;	Length	DB 106;	Score	DB 107;	Length	DB 108;	Score	DB 109;	Length	DB 110;	Score	DB 111;	Length	DB 112;	Score	DB 113;	Length	DB 114;	Score	DB 115;	Length	DB 116;	Score	DB 117;	Length	DB 118;	Score	DB 119;	Length	DB 120;	Score	DB 121;	Length	DB 122;	Score	DB 123;	Length	DB 124;	Score	DB 125;	Length	DB 126;	Score	DB 127;	Length	DB 128;	Score	DB 129;	Length	DB 130;	Score	DB 131;	Length	DB 132;	Score	DB 133;	Length	DB 134;	Score	DB 135;	Length	DB 136;	Score	DB 137;	Length	DB 138;	Score	DB 139;	Length	DB 140;	Score	DB 141;	Length	DB 142;	Score	DB 143;	Length	DB 144;	Score	DB 145;	Length	DB 146;	Score	DB 147;	Length	DB 148;	Score	DB 149;	Length	DB 150;	Score	DB 151;	Length	DB 152;	Score	DB 153;	Length	DB 154;	Score	DB 155;	Length	DB 156;	Score	DB 157;	Length	DB 158;	Score	DB 159;	Length	DB 160;	Score	DB 161;	Length	DB 162;	Score	DB 163;	Length	DB 164;	Score	DB 165;	Length	DB 166;	Score	DB 167;	Length	DB 168;	Score	DB 169;	Length	DB 170;	Score	DB 171;	Length	DB 172;	Score	DB 173;	Length	DB 174;	Score	DB 175;	Length	DB 176;	Score	DB 177;	Length	DB 178;	Score	DB 179;	Length	DB 180;	Score	DB 181;	Length	DB 182;	Score	DB 183;	Length	DB 184;	Score	DB 185;	Length	DB 186;	Score	DB 187;	Length	DB 188;	Score	DB 189;	Length	DB 190;	Score	DB 191;	Length	DB 192;	Score	DB 193;	Length	DB 194;	Score	DB 195;	Length	DB 196;	Score	DB 197;	Length	DB 198;	Score	DB 199;	Length	DB 200;	Score	DB 201;	Length	DB 202;	Score	DB 203;	Length	DB 204;	Score	DB 205;	Length	DB 206;	Score	DB 207;	Length	DB 208;	Score	DB 209;	Length	DB 210;	Score	DB 211;	Length	DB 212;	Score	DB 213;	Length	DB 214;	Score	DB 215;	Length	DB 216;	Score	DB 217;	Length	DB 218;	Score	DB 219;	Length	DB 220;	Score	DB 221;	Length	DB 222;	Score	DB 223;	Length	DB 224;	Score	DB 225;	Length	DB 226;	Score	DB 227;	Length	DB 228;	Score	DB 229;	Length	DB 230;	Score	DB 231;	Length	DB 232;
-------------	-----------------------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------

RESULT 8	G6PI_YERPE	STANDARD;	PRT;	548 AA.
ID	G6PI_YERPE			
AC	Q8ZAS2;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose			
DE	isomerase) (PGI) (Phosphohexose isomerase) (PHI).			
GN	PGI OR YPO3718.			
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Yersinia.			
NCBI_TaxID=632;				

```

RN [1] SEQUENCE FROM N.A.
RP STRAIN-CO-92 / Biovar Orientalis;
RC MEDLINE=21470413; PubMed=11586360;
RX parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebathia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -I- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
CC phosphate.
CC -I- PATHWAY: Involved in glycolysis and in gluconeogenesis.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE GPI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AJ141458; CAC93186.1; -.
CC DR InterPro: IPR001672; G6P_Isomerase.
CC DR Pfam: PF00342; PGI; 1.
CC DR PRINTS: PR00662; G6PIISOMERASE.
CC DR PROSITE: PS00765; P_GLUCOSE_ISOMERASE_1; 1.
CC DR PROSITE: PS00174; P_GLUCOSE_ISOMERASE_2; 1.
CC KW Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
CC FT ACT_SITE 386 386 BY SIMILARITY.
CC FT ACT_SITE 514 514 BY SIMILARITY.
CC FT SEQUENCE 548 AA; 61161 MW; 3ED9D49B3A308A3B CRC64;
SQ
Query Match 5.4%; Score 103; DB 1; length 548;
Best Local Similarity 21.1%; Pred. No. 0.85;
Matches 75; Conservative 57; Mismatches 106; Indels 118; Gaps 20;
QY 56 RALNSYFPPVEESALERPETIS-----EKTYYDLTNEETDSTTS 98
DB :||:||| :||| :||| :||| :||| :||| :|||
DB 13 KALOQHFE-----QMKVITISLSFAKDDQRFNRSATFDQMLVDFSKNRITSETLE 64
QY 99 KISPSEDTOENG-----SMESLITWNIDGLDNLSEARGVCSYLAL---YSPDVI 148
DB ||:|||:||| ||||| |:|||:||| :||| :||| :||| :||| :|||
DB 65 KL---QDLAKETDLGAIKSMFS-----GEKINRTEDRA--VLHIALRNSNTPIVV 111
QY 149 FLQEVIPPYSYLKRRSSNYE-IITGHEGYFTAIMLKSRVKLSQEIIPEPSTKMNR- 206
DB ::||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 112 DGKDVMPEVNAVLAKKMKQFCDRYISGDWKGYTGKAITDVVNIGIGSDLGPMYTEALRP 171
QY 207 --NLICVHVNVSGNELCLMTSHLESTRGHAERMQLK-----MVLKKMQEAPEASAT-- 256
DB |:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 172 YKNHLNMH-----EVSNDVGT-HIAEALKPLNPETTLFLVASKTFTTQETMTNA 219
QY 257 -----VIFAGD-----TNLRDREVTTCGGLPNNIVDWEEFLGPKPHCYT- 296
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 220 HSARDWFLSAAGDPAHVAKHFAALSTNA-KAVGEFGIDTNMMFEFWDWVG---RYSYL 273
QY 297 WDTQMSNLGITAACTLR-----DRIFFRAAEFGHIIPRSLDLLGL 339
DB |:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 274 W-----SAIGLSIALSVGFHFEGQLLSGAHAMDKHFAETPRAEKX--LPVLLALLIGI 322
RESULT 9
EX3_HAEIN STANDARD; PRT; 267 AA.
AC P44318;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
```

```

DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Exodeoxyribonuclease III (EC 3.1.11.2) (Exonuclease III) (EXO III).
OS XTHA OR HI0041.
SN Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
ON NCBI_TaxID=727;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uutterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: MAJOR APURINIC-APYRIMIDINIC ENDONUCLEASE OF E.COLI. IT
CC REMOVES THE DAMAGED DNA AT CYTOSINES AND GUANINES BY CLEAVING ON
CC THE 3' SIDE OF THE AP SITE BY A BETA-ELIMINATION REACTION
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Degradation of double-stranded DNA. It acts
CC progressively in a 3'-to 5'-direction, releasing nucleoside 5'-
CC phosphates.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AP/EXO A FAMILY OF DNA REPAIR ENZYMES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: U32689; AAC21719.1; .
DR HSSP: P09030; IAKO.
DR TIGR: HI0041; .
DR InterPro: IPR000097; Apendonclsel.
DR InterPro: IPR004442; ExoDNase_III.
DR InterPro: IPR004808; ExoIII_xth.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR TIGRFAMS: TIGR00195; exoDNase_III; 1.
DR TIGRFAMS: TIGR00633; xth; 1.
DR PROSITE: PS00726; AP_NUCLEASE_F1_1; 1.
DR PROSITE: PS00727; AP_NUCLEASE_F1_2; 1.
DR PROSITE: PS00728; AP_NUCLEASE_F1_3; 1.
DR Hydrolase; Nuclease; Exonuclease; DNA repair; Complete proteome.
KW METAL 34 34 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT ACT_SITE 259 259 GENERAL BASE (BY SIMILARITY).
FT SEQUENCE 267 AA; 31025 MW; 6B3ADE465A1E347C CRC64;
SQ
Query Match 5.4%; Score 102.5; DB 1; Length 267;
Best local Similarity 21.1%; Pred. No. 0.34;
Matches 59; Conservative 45; Mismatches 108; Indels 67; Gaps 13;
QY 117 ITWNIDGLDNLNUSERARGVCSYLALSPDVIFLQE--VIPPYSYLKKRSSNYEITGH 174
DB 4 ISFNING-----LRARPHQLEAIIETKQPDYVIGLQEIKADEAFPEYTEITENLGHYHFHGG 58
QY 175 EEGYFTAIMLKSRVKLSQETIPEPSTKMNRNLGVHNVNSGNELCLMT----- 224
DB 59 QKGHYGVALLTQEPKV-IRRGFPTDNEDAQKRIIMADLE---TEFGLLTVINGYFPGGE 114
QY 225 SHLESTRGHAERM-NOLKMKVLKKMOEAPESATVIFAGDTNLDREVTF----- 271

```

Db 115 SRAHETKPAKEKYADLQOYLEK--EHDKSNPILIMGDMNISPSDLDIGIDENRKRWL 172
QY 272 ---RCGGLPN-----NIVDWEEFLGPKHCQYTW----DTQMSNLGITACKL 313
Db 173 RTGKCSFLPEERAWYQRLYDYGLEDSFRKLNPANDKFSWFDYRSKGFDDNRG-----L 226
QY 314 RFDRIFR-RAAAEGHIIPRSLDLGLEKDCGRFSDH 351
Db 227 RIDHILVSQKLAERCVDVGIALDIRAMEK-----PSDH 259

RESULT 10

G6PI_SALTY
ID G6PI_SALTY STANDARD; PRT; 549 AA.
AC Q82107;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose
isomerase) (PGI) (Phosphohexose isomerase) (PHI).
GN PGI OR STY4417.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
Feltwell T., Hamlin N., Haque A., Hlen T.T., Holroyd S., Jagels K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
phosphate.
CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL627282; CAD09205.1; -.
DR InterPro; IPR001672; G6P_Isomerase.
DR Pfam; PF00342; PGI; 1.
DR PRINTS; PR00662; G6PISOMERASE.
DR PROSITE; PS00765; P_GLUCOSE_ISOMERASE_1; 1.
DR PROSITE; PS00174; P_GLUCOSE_ISOMERASE_2; 1.
KW Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
FT ACT_SITE 386 BY SIMILARITY.
FT ACT_SITE 514 BY SIMILARITY.
SQ SEQUENCE 549 AA; 61414 MW; 3CFE39008185D56E CRC64;

Query Match 5.4%; Score 102.5; DB 1; Length 549;
Best Local Similarity 21.3%; Pred. No. 0.94;
Matches 67; Conservative 55; Mismatches 90; Indels 103; Gaps 18;

OY 85 VDLTNEETDSTTSKISPSSEDTQENG-----SMFSLTWNIDGLDLNLSERARGVCS 138
Db 51 VDFSKNRITFEETLAKL---QDLAKETPLDLAGAIKSMFS-----GEKINRTEDRA---VL 97

QY 139 YLAL-----YSPDIVFLQEVIPPYSYLKK-RSSNYEITTGHEGYFTAIMLKSRVKLS 193
Db 98 HVALRNSNTPIIIVDGKDVMEVNAVLEKMKTFSSQAIISGQWKGYTGKAITDVYVNIIGIG 157
QY 194 QEIIPEPSTKMR---NLGVYVNVSGNELCLMTSHLESTRGHAERMNOLKMYLKKMOE 250
Db 158 SDLGPEWVTALRPYKNHLMH-----FVSNVDGT-HIAE-----VLKKVN- 197
QY 251 APESATVIFAGDT-----NLRD-----REYTRCGGL 276
Db 198 -PETTLELVASKTFTTQETMTNAHSARDWFLKSAGDEKHAVAKHFAALSTNAKAVGEFGID 256
QY 277 PNNIVDWEEFLGPKHCQYT-WDTQMSNLGITACKLRFD---RIFFRAAEEGHI--- 329
Db 257 TANMEFEWDMVVG---RYSIW-----SAIGLSIISVGFDFNVELLSGAHAMDKHFSTT 307
QY 330 -----IPRSLDLGL 339
Db 308 PAEKNLPIILLALIGI 322

RESULT 11

G6PI_SALTY
ID G6PI_SALTY STANDARD; PRT; 549 AA.
AC Q82K14;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose
isomerase) (PGI) (Phosphohexose isomerase) (PHI).
GN PGI OR STM4221.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
phosphate.
CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE008897; AAL23045.1; -.
DR StGene; SG?????; pgi.
DR InterPro; IPR001672; G6P_Isomerase.
DR Pfam; PF00342; PGI; 1.
DR PRINTS; PR00662; G6PISOMERASE.
DR PROSITE; PS00765; P_GLUCOSE_ISOMERASE_1; 1.
DR PROSITE; PS00174; P_GLUCOSE_ISOMERASE_2; 1.
KW Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
FT ACT_SITE 386 BY SIMILARITY.
FT ACT_SITE 514 BY SIMILARITY.
SQ SEQUENCE 514 AA; 514 MW; 3CFE39008185D56E CRC64;

[illegible]

DR InterPro; IPR001572; RNA_pol_B.
DR Pfam; PF00562; RNA_pol_B: 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KW Transferrase; Transcription; DNA-directed RNA polymerase;
Complete proteome.
SQ SEQUENCE 1391 AA; 155621 MW; B2F345AB24F18EAD CRC64;

Query Match 5.4%; Score 102; DB 1; Length 1391;
Best Local Similarity 22.1%; Pred. No. 3.8;
Matches 85; Conservative 64; Mismatches 115; Indels 120; Gaps 24;

```

QY      44 QCFLENDWEMERALNSYPPEVE-----ESALER-RPETISEPKT----- 83
           ||| | : : | : || : : : : : : : : : : : : : : : : : : : :
Db      42 QTFL--DHDLENLIAAYF--PIKSPNDRYTINFKGLRRTAPERNEAQSRSSESKTYETGI 96

QY      84 YVDLTNETTTSDSTTSKISPSEDTQENGSMFSLITWNIDGDLNNLSERAGVCSYLALY 143
           ||| | | | | | | : : : : : : : : : : : : : : : : : : : :
Db      97 YADL--ELIDSATGTIKRKPSKKNSA-----TSSVDGVFLT-----LPLI 136

QY     144 SPDVIQLQEVY-----PPYY---SYLKKRSSNYELITGHEGYFT-----A 181
           : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     137 TRDGVFIIVNGIEKEFVIAQTTRSPGIYMLTKSOLKLSSSRKV----QEGYVCVLPANGS 192

QY     182 IML----KKSrvKLKSQEII-----PFPSTKMRRNLICVHNVSGNELC----- 221
           : || : : : : : : : : : : : : : : : : : : : : : : : : :
Db     193 VMLIYISNKKKIEDIAFVQILLRDVAREGAKIFPITTLKAF-----GMSGKEILKVFKN 247

QY     222 -LMTSHLESTRGHAAERMN----QLKVLKKMQEAPESATVIFAGDTNLRDREVTTCGG 276
           : | || : : : : : : : : : : : : : : : : : : : : : : : :
Db     248 EFITRSLAEAVYNAKDPLNNVDPEIKNLLREFRD-----GKTDLRRKGIASDQKI 297

QY     277 PNNIVDVWEFLGPKKHCOYWTDMQNSNLG-ITACKLRFDRIFFRAAEEGHIIIPR-SL 334
           : : | : | : : : : : : : : : : : : : : : : : : : : : : :
Db     298 RSIYS DY--VLLKEHKALSEAKPNDPKVGQLEADMDELMDKIITERAAK--HIVHEL 353

QY     335 DLGLEKLD-CGRFPSDHWGILCN 357
           | ||| | | : : : ||| :
Db     354 SLRGLENTDECPE--NSYHALICS 375
```

RESULT 13
PHL1_BACCE
ID PHL1_BACCE STANDARD; PRT; 333 AA.

```

AC P09599;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sphingomyelinase C precursor (EC 3.1.4.12) (Sphingomyelin
   phosphodiesterase) (Smase) (SMPLC) (Cereolysin B).
GN SPH.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE-1;
RX MEDLINE=89057484; PubMed=2848222;
RA Johansen T., Haugl F.B., Ikezawa H., Little C.;
RT "Bacillus cereus strain SE-1: nucleotide sequence of the
   sphingomyelinase C gene.";
RL Nucleic Acids Res. 16:10370-10370(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VKM 164;
RX MEDLINE=93249510; PubMed=8387306;
RA Kuzmin N.P., Gavrilenko I.V., Krukov V.M., Karpov A.V.;
RT "Nucleotide sequence of phospholipase C and sphingomyelinase genes
   from Bacillus cereus BKM-B164 (letter).";
RL Bioorg. Khim. 19:133-138(1993).
RN [3]
RP SEQUENCE OF 1-188 FROM N.A.
RC STRAIN=SE-1;
RX MEDLINE=88313678; PubMed=3137122;
```


AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM V0).
RX MEDLINE=95105188; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT "Characterization of the complete genomic structure of the human
RT versican gene and functional analysis of its promoter.";
RL J. Biol. Chem. 269:32999-33008(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V1).
RC TISSUE=Placenta;
RX MEDLINE=90059882; PubMed=2583089;
RA Zimmermann D.R., Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican.";
RL EMO J. 8:2975-2981(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM V2).
RC TISSUE=Glial tumor;
RX MEDLINE=95105187; PubMed=7806529;
RA Dours-Zimmermann M.T., Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RT alternative splice variants of human versican.";
RL J. Biol. Chem. 269:32992-32998(1994).
RN [4]
RP SEQUENCE OF 2711-3396 FROM N.A.
RC TISSUE=Lung fibroblast;
RX MEDLINE=88007514; PubMed=2820964;
RA Krusius T., Gehlsen K.R., Ruoslahti E.;
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RT lectin-like and growth factor-like sequences.";
RL J. Biol. Chem. 262:13120-13125(1987).
RN [5]
RP SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE=93122792; PubMed=1478664;
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA McPherson J.D.;
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RT human chromosome 5 (5q12-5q14).";
RL Genomics 14:845-851(1992).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RC TISSUE=Brain;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [7]
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RC TISSUE=Aortic smooth muscle;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [8]
RP PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial

RT hyaluronate-binding protein.";
RL J. Biol. Chem. 264:5981-5987(1989).
RN [9]
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors.";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
CC V2, V3 and Vint; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
CC in normal brain, gliomas, medulloblastomas, schwannomas,
CC neurofibromas, and meningiomas; V2 is restricted to normal brain
CC and gliomas; V3 is found in all these tissues except
CC medulloblastomas.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U16306; AAA65018.1; -.
DR EMBL; X15998; CAA34128.1; -.
DR EMBL; S52488; AAB24878.1; -.
DR EMBL; U26555; AAA67565.1; -.
DR EMBL; D32039; BAA06801.1; -.
DR EMBL; J02814; AAA36437.1; -.
DR EMBL; AF084545; AAD48545.1; -.
DR PIR; S06014; S06014.
DR PIR; A29348; A29348.
DR PIR; A30358; A30358.
DR HSSP; P01132; 1EGF.
DR Genew; HGNC:2464; CSPG2.
DR MIM; 118661; -.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBLDOD.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR	PROSITE; PS00022; EGF_1; 2.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01187; EGF_CA; 1.
DR	PROSITE; PS01241; LINK; 2.
DR	PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR	PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW	Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW	Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KM	Hyaluronic acid; Alternative splicing.
FT	SIGNAL. 1 20 POTENTIAL.
FT	CHAIN 21 3396 VERSICAN CORE PROTEIN.
FT	DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
FT	DOMAIN 167 244 LINK 1.
FT	DOMAIN 265 346 LINK 2.
FT	DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
FT	DOMAIN 1336 3089 GAG-BETA.
FT	DOMAIN 3089 3125 EGF-LIKE 1. .
FT	DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL) .
FT	DOMAIN 3176 3290 C-TYPE LECTIN.
FT	DOMAIN 3295 3353 SUSHI.
FT	DISULFID 44 130 BY SIMILARITY.
FT	DISULFID 172 243 BY SIMILARITY.
FT	DISULFID 196 217 BY SIMILARITY.
FT	DISULFID 270 345 BY SIMILARITY.
FT	DISULFID 294 315 BY SIMILARITY.
FT	DISULFID 3093 3104 BY SIMILARITY.
FT	DISULFID 3098 3113 BY SIMILARITY.
FT	DISULFID 3115 3124 BY SIMILARITY.
FT	DISULFID 3131 3142 BY SIMILARITY.
FT	DISULFID 3136 3151 BY SIMILARITY.
FT	DISULFID 3153 3162 BY SIMILARITY.
FT	DISULFID 3169 3180 BY SIMILARITY.
FT	DISULFID 3197 3289 BY SIMILARITY.
FT	DISULFID 3265 3281 BY SIMILARITY.
FT	DISULFID 3296 3339 BY SIMILARITY.
FT	DISULFID 3325 3352 BY SIMILARITY.
FT	CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 330 330 N-LINKED (GLCNAC. . .) .
FT	CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 809 809 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 1442 1442 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 1468 1468 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 1663 1663 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 2179 2179 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 2280 2280 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 2360 2360 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL) .

Query Match	5.28;	Score 98.5;	DB 1;	Length 3396;
Best Local Similarity	22.08;	Pred. No. 26;		
Matches	76;	Conservative	53;	Mismatches 131;
				Indels 85;
				Gaps 18;

QY	7	LEGGRAAEEEGEPEVKKRRLLCVEE-----ASVASCDA--VAQCFLAENDWEMERAL	58
	:	:	: : : :
Db	2066	VEGTRKAPVEKE--EVKVSGTVSTNFPQTIEPAKLWSRQEVNPVRQEISETTSEEQIQE	2122
QY	59	NSYFEPVVEESALERRPBTISEPKTYVDLTJNEETD---STTSKISPSEDTOQENGSMF	114
	:	:	: : : :
Db	2123	EKSFESEPONSATE---QTIFDSOTFTE-TELKTIDYSVLTTKKTYSDDKEMEDTSLV	2178
QY	115	SLITWNIDGLDLNNLSERARGVCYSYLALYSPD-----VIFLQEVIPYY----SY	160
	:	: :	: : :
Db	2179	NMSTPDPP-----ANGLESYTTL-PEATREKSHFFLATALVTESISPAEHVYTDSP	2226
QY	161	LKRSSNYEIIITGHEGYFTAIMLKSRVKL-----KSOEIFP-----FPSTKMMR	206
	:	:	: : : :
Db	2227	IKKEESTKHFPKGMRP-----TIQESDTELLESGIGSEEVLPPLLPTLESVNFTEVEQIN	2280

QY	207	NILCVHVNVSNGNELCLMTSHLESTRGHAERMNQLKMYLKKMQEAP-ESATVIFAGDTNL	265
Db	2281	NELYPH-----TSQVESTSSDKIEDFNRMENVAKEV--GPLVSQTDFEGSGSV	2327
QY	266	RDR---EVTRCGGLPNNIVDWEF--LGKPKHCQYTWTQMNSN	304
Db	2328	TSTTLIEILSDTGAEGPTVAPLPFSTDIGHPONQTVRWAAEEIQT	2372

Search completed: November 14, 2002, 10:25:45
Job time : 12.4003 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 ; Search time 13.5673 Seconds
(without alignments)
2565.041 Million cell updates/sec

Title: US-09-697-863A-2
Perfect score: 1902
Sequence: 1 MELGSCLEGGREAAEEEGEP.....DCGRFPSDHWGLICNLDIIL 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	420	22.1	362	2	T27272	hypothetical prote
2	231	12.1	441	2	B86252	hypothetical prote
3	113	5.9	502	2	T40792	hypothetical prote
4	111.5	5.9	252	2	D95853	conserved hypothet
5	110.5	5.8	361	2	S68268	apurinic/apurimidi
6	108	5.7	253	2	T35506	hypothetical prote
7	108	5.7	477	2	S50477	hypothetical prote
8	105.5	5.5	253	2	D90737	hypothetical prote
9	105.5	5.5	253	2	E85587	hypothetical prote
10	105.5	5.5	253	2	F64815	yhbp protein - Esc
11	105.5	5.5	673	2	T46340	hypothetical prote
12	105	5.5	1374	2	A84888	hypothetical prote
13	104.5	5.5	281	2	AG2333	hypothetical prote
14	103.5	5.4	256	2	AD1115	exonuclease III ho
15	103	5.4	288	2	T24066	exonuclease III ho
16	103	5.4	548	2	AF0452	glucose-6-phosphat
17	102.5	5.4	267	2	H64044	exodeoxyribonuclea
18	102.5	5.4	549	2	AD1013	glucose-6-phosphat
19	102	5.4	265	2	F97566	probable nuclease
20	102	5.4	265	2	AD2787	exonuclease III li
21	102	5.4	1391	2	S73652	RNA polymerase bet
22	101	5.3	1275	2	I38588	reverse transcript
23	101	5.3	1275	2	S65824	reverse transcript
24	100	5.3	1261	2	E59430	PTPL1-associated R
25	99.5	5.2	333	2	S01950	sphingomyelin phos
26	99.5	5.2	592	2	PS0197	phospholipase C (E
27	98.5	5.2	563	2	F64130	glucose-6-phosphat
28	98.5	5.2	2409	1	A60979	versican precursor
29	98	5.2	3844	2	T18402	asparagine/asparta

30	97.5	5.1	1275	2	B28096	line-1 protein ORF
31	97	5.1	270	2	H83327	exodeoxyribonuclea
32	96.5	5.1	658	2	T33568	hypothetical prote
33	96.5	5.1	996	2	S46812	hypothetical prote
34	96.5	5.1	3191	2	T22945	hypothetical prote
35	96	5.0	289	2	AG3260	exodeoxyribonuclea
36	95	5.0	267	2	C97690	exodeoxyribonuclea
37	95	5.0	267	2	AH2915	exodeoxyribonuclea
38	95	5.0	1280	2	B34087	hypothetical prote
39	94.5	5.0	547	2	T30704	rifampicin resista
40	94	4.9	255	2	E70166	exodeoxyribonuclea
41	93.5	4.9	295	2	S34544	hypothetical prote
42	93.5	4.9	388	2	T47530	hypothetical prote
43	93.5	4.9	892	2	T50985	related to transcr
44	93	4.9	549	1	NUEC	glucose-6-phosphat
45	93	4.9	549	2	H91254	glucosephosphate 1

ALIGNMENTS

RESULT 1
T27272
hypothetical protein Y63D3A.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27272
R:White, S.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20335
A:Accession: T27272
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-362 <WIL>
A:Cross-references: EMBL:AL032652; PIDN:CAA21707.1; GSPDB:GN00019; CESP:Y63D3A.4
A:Experimental source: clone Y63D3A
C:Genetics:
A:Gene: CESP:Y63D3A.4
A:Map position: 1
A:Introns: 18/1; 43/1; 116/3; 148/2; 225/3; 320/2

Query Match	22.1%	Score 420;	DB 2;	Length 362;
Best Local Similarity	29.3%;	Pred. No. 2.4e-26;		
Matches 108;	Conservative 75;	Mismatches 142;	Indels 44;	Gaps 10;
QY	7	LEGGREAAEE-----GEPEYKKRRLLCV-----EFASVASCDAVAOCLAEN 50		
DB	11	IEAKRQKMSQEDSEVEIEILDEPEQGLKNSSMSDEQKLHFAIITATDEAFAQSIIDV 70		
QY	51	DWEMERALNSYFEPPEVESALERPPTISEPKTYVDLTNETTDSITSKISPSEDTQEN 110		
DB	71	DWDLKKALDVFY-----GSEAFPAEARSAAVMGASSMASSGAAVMTAEDLK--- 116		
QY	111	GSMFLITWNIDGLDLNNLSERARGVCSYLALYSPDVIFLOEVIPIPYSYLKKRSSNVEI 170		
DB	117	GFEVSVMWNIDGLDGRSLTFRMKAVAHIVKNVNPDLFLQEVVDRDLAPIDKLQSLYKI 176		
QY	171	ITGHEE-GYFTAIMLKSRVKLSQELIPFPSTKMMRNLLCVHNVSGNELCLMTSHLES 229		
DB	177	YYSNKGCCQYTAIIVSK-MFDVEKHVDYIHFNQSGMYRTLQILEGSIGGLKVFLLNTHLES 235		
QY	230	TRGHAERKMNLKMYLKKMQE---APESATVIFAGDTNLRDREYTRCGGLPNNIVDVEEF 286		
DB	236	TREHRPQRCAGFGFCMDKVREIIAQNPCALVFFGGDLNLRDDEVS---VPDGVKDAWEA 292		
QY	287	LGPKPHCQYTWDTQMSNLSGITAACKLRFDRIFFRAAAECHITPRSLDLGLEKL-DCG 345		
DB	293	AGSDNKTKTFTWDTFFKNDNKQGFHGAKMRFDRLYW-----SGPLDKVKFTLEGRQIRSL 347		
QY	346	RFPSDHWGL 354		
DB	348	CFPSDHWAI 356		

QY 136 VCSYLALSPDVIFLQEV-----IPPYSYLKKRSSNVEIITGHEEGYFTAM 183
: : : | | | | | : : : : : : : : :
Db 34 IAAVIAECQPDVIALQEV DVGARTGIDQAHMIATHLNMEAEFHPALHLEDEKYGDAVL 93
QY 184 LKKSRLVKLSQEIIPFPSTKMMRNLLCVHVNVSNGNELCLMTSHLESTRGHAERMNQLKM 243
: : : : : : : : : : : : : : : : : : :
Db 94 ---TALPMRLIKAAPLPSSESRGALMWEIDVAVKLQIVFTHL-GLRG--AERLRQATA 147
QY 244 VLKK---MQEAPESATVIFAGDTNLRDREYTRCGGLPNNTVDVMEFLG-KPKHCQYTWDT 299
: : : | | | | | : : : | | : : : : :
Db 148 LLPGWLGMAQGDAAHVVLAGDLNATGRS-TAYRLLARQLSDAQLLTGVKPR-----P 199
QY 300 QMNSNLGITACKLRFDRIFFRAAAEHGHIIPRSLDLGLEKLDG-----GRFPSDHW 352
: : : | | : : : : : : : : : : : : :
Db 200 TFP SRLPL-----LRID-----HVLVGK----GIEVASCRVHGSTLARSASDHL 239
QY 353 GLLCNLDTIL 362
: : : | | : : : : : : : : : : : : :
Db 240 PLLAELDVVM 249

RESULT 5
S68268
apurinic/apurimidinic(AP)-endonuclease class II - slime mold (Dictyostelium discoideum)
C/Species: Dictyostelium discoideum
C/Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 18-Feb-2000
C/Accession: S68268
R/Freeeland, T.M.; Guyer, R.B.; Lin, A.Z.; Deering, R.A.
Nucleic Acids Res. 24, 1950-1953, 1996
A/Title: Apurinic/apurimidinic (AP)-endonuclease from Dictyostelium discoideum: cloning,
A/Reference number: S68268; MUID:96226184; PMID:8657579
A/Accession: S68268
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-361 <PRE>
A/Cross-references: EMBL:U31631; NID:g1215675; PIDN:AAC47024.1; PID:g967268
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1995
C/Superfamily: exodeoxyribonuclease III

Query Match 5.8%; Score 110.5; DB 2; Length 361;
Best Local Similarity 21.5%; Pred. No. 0.31;
Matches 62; Conservative 52; Mismatches 94; Indels 81; Gaps 15;
QY 8 EGGREAAEEGEPEVKKRRLLCVERASVASCDAAVAQCFLAENDWEMERALNSYFEPVE 67
| : | | | | | | : |
Db 27 EEEKEVEEEE-EEDKKRKLKTPPAKKA-----PAAK 57
QY 68 ESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOE---NGSMFSLITWNIDGL 124
: : : : : : : : : : : : : : : : : : :
Db 58 KAAAKKKSK--DEDEDEKEEKEETNKTTASVSIAIDNLEPKVEENQMKIISWVAGF 115
QY 125 DLNLSERAGVCSYLALSPDVIFLQEV-IPPYSYLKK---RSSNYEITGHEEGYF 179
| : : | : : : : : : : : : : : : :
Db 116 K---SVLSKGFTECEVENPDVLCQETKINP--SNKKQMPKGYEHFIADQGH 169
QY 180 -TAIMLKRSRVKLSQEIIPFPSTKMMRNLLCVHVNVSNGNELCL-----MTSHLES-- 229
: : : | : : : : : : : : : : : : :
Db 170 GTGVLTKKK-----PNAITFGIGIAKHND-EGRVITLLEYVQFYIVNTYIPNAG 216
QY 230 TRGHAERMNQLKVLK-----MQEAPESATVIFAGDTNLRDREV 270
: : : : : : : : : : : : : : : : : : :
Db 217 TRG-----LQRLDYRIKEWDVDFQAYLEKLNATKPIIWCGLNVAHTEI 260

RESULT 6
T35506
hypothetical protein SC6E10.19c - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C/Accession: T35506
R/Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A/Reference number: Z21580

A/Accession: T35506
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-253 <SEE>
A/Cross-references: EMBL:AL109661; PIDN:CAB51973.1; GSPDB:GN00070; SCOEDB:SC6E10.19c
A/Experimental source: strain A3(2)
C/Genetics:
A/Gene: SCOEDB:SC6E10.19c
Query Match 5.7%; Score 108; DB 2; Length 253;
Best Local Similarity 19.5%; Pred. No. 0.3;
Matches 56; Conservative 52; Mismatches 119; Indels 60; Gaps 12;
QY 95 STTSKISPSEDTQQENGSMFSLITWNIDGLNLSERAGVCSYLALSPDVIFLQEV 154
: : : | : : : : : : : : : : : : :
Db 2 ATSLPALPNSRTEPDGSAVIRVLSYNI-----RSLRDDTDALARVIAACAPDLVLLQEA- 55
QY 155 PPYSYLKK-----RSSNYEITGHEEGYFTAIMLK-KSRVKLSQEIIPFPSTKMMRN 208
: : : | : : : : : : : : : : : : :
Db 56 PFRFRWKKITRLAAGDLVLLSGGTAAGPALCSLRATVERTEDVLLPLTPGRHRRGI 115
QY 209 LCVHVNVSNGNELCLMTSHLESTRGHAERMNQLKVLKMQEAPESATVIFAGDTNLR-- 266
: : : | : : : : : : : : : : : : :
Db 116 AAAVRIRIGARIGVLSHLSL--DADERHEQAGLLDLH-LAALGVKAAVAGD LNERPG 171
QY 267 DREVTRCGLPNNTVDVMEFLGPKHCQYTWDTQMSNLGITACKLRFDRIFFRAAEE 326
: : : | : : : : : : : : : : : : :
Db 172 GRTERRLG---EGLRDCW--TAAPWGGEYTFP-----ATAPDRIDAVFVTE---- 213
QY 327 GHIIPRSLDLGLEKLDG-----RFPDHWGLLCNLDI 360
: : : | : : : : : : : : : : : : :
Db 214 -----GIEVLGCGVPGLAGVAEDDLRAATDHLPVLTALRV 249

RESULT 7
S50477
hypothetical protein YER019w - yeast (Saccharomyces cerevisiae)
C/Species: Saccharomyces cerevisiae
C/Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
C/Accession: S50477
R/Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A/Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lamb
A/Reference number: S50433
A/Accession: S50477
A/Molecule type: DNA
A/Residues: 1-477 <DIE>
A/Cross-references: EMBL:U18778; NID:g603592; PID:g603611; GSPDB:GN00005; MIPS:YER019
A/Gene: SGD:ISCL; MIPS:YER019w
A/Cross-references: SGD:S0000821
A/Map position: 5R

Query Match 5.7%; Score 108; DB 2; Length 477;
Best Local Similarity 22.0%; Pred. No. 0.74;
Matches 78; Conservative 47; Mismatches 124; Indels 106; Gaps 20;
QY 90 EETDSTTSKISPSEDTQQENGSMFSL-ITWNIDGLDL--NLSERAGVCSYLALYSP 145
: : | : : : : : : : : : : : : :
Db 13 KEDGQSEFEALNGTNAIMSDNSKAYSIKELTFNTWGLKYVSKHRKRLRATADKLACHSM 72
QY 146 -----DVIFLQEV-IPPYSYLKKR-SSNYEITGHEEGYFT-- 180
: : : | : : : : : : : : : : : : :
Db 73 LTPISDELLPNGSDSNENEDYDIALQELIWCVEDWKYLASACASKYQRLFHSGILTGP 132
QY 181 --AIMLKSRVKLSQEIIFP-----STKMMRNLLCVHVNVSNGNELCLMTSHL 227
: : : | : : : : : : : : : : : : :
Db 133 GLAIL---SKVPIESTFLYRFPINGRPSAVFRGDWYVGKSIATYVLTGTPIAIMNSHM 189
QY 228 E---STRGHAERMNQ-----LKMVLKMQEAPESATVIFAGDTNLRDREVTTCGGLPN 278
: : : | : : : : : : : : : : : : :
Db 190 HAPYAKQGDAAVLCRSCQAWDFSLIKLYRQA--GYAVIVGDLN-----SRPGSLPH 241

```
QY 279 -----NIVDWEFL-GKP-----KHQÇ-----YTWDTQMNSNLGITA 309
      :||| |||
Db 242 KFLTQEAGLVDSEMEQLHGKODLAVIARLSPLQOLLKGCCTCDSLNTWTWRAQRPD----E 297
      :||| |||
QY 310 ACKLRF--DRIFERRAAAEEGHII PRSDLLGLEKLDCGRFPSPDHWGILCNDIT 361
      ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 ACRLDYALIDPDLQTVDAGVRFTR-----IPHLDCS--VSDHFAYSCLTINIV 344
```

RESULT 8
D90737
hypothetical protein ECs0868 [Imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90737
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
Data Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAB34291.1; PID:g13360327; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs0868

	Query Match	5.58;	Score 105.5;	DB 2;	Length 253;	
	Best Local Similarity	20.8%;	Pred. No. 0.47;			
	Matches 59;	Conservative 33;	Mismatches 82;	Indels 109;	Gaps 11;	
QY	104 EDTQOENGSMFSLITWNID-----GLDLNLTSEBARGVCSYLALYSPPVIFLQEV--	153	:	:	::	::
Db	3 DQTQOFS--FKVLTIINHKGFTAFNRREFILPELRDAVRTV-----SADIVCLOEVMG	52				
QY	154 ---IPPYSYLLKRSNVEIITG-----HEEGYFTAIMLKSRVKIKSQE--	195	:	:	:: :	: :
Db	53 AHEVHPLHVENPDTSHYEFLADTMSDFAYGRNAVYPFGHHGNAVL--SRYPPIEHENR	110				
QY	196 -----IIEPSTKMRRNLCLGVHNVSGNELCLMTSHLESTRGHAAERMN	239	:	:	:: :	:
Db	111 DVSVDGAEKRGVLYCRIVP-PMTGKAIHVMCVHLGL-----REAHRQA	152				
QY	240 QLKMLVKKMQEAPESATVIFAGDTNLRDREVTRCGGLPNNIIVDWEFELGPKPHCQYTWD	299	:	:	:: :	:
Db	153 QLAMLAEMVNELPDGEVYLVAADF-----DWKQ	181				
QY	300 QMNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLLGLEKL	342	:	:	:	:
Db	182 KANHPLKVQAG----LDEIFTTRAHGPRARTFPVOEPILLRDRI	220				

RESULT 9
E85587
hypothetical protein ybhp [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85587
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85587
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <STO>
A:Cross-references: GB:AE005174; NID:g12513779; PIDN:AAG55161.1; GSPDB:GN00145; UWGP:Z10
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ybhp

Query Match	5.5%;	Score 105.5;	DB 2;	Length 253;
Best Local Similarity	20.8%;	Pred. No. 0.47;		
Matches 59;	Conservative 33;	Mismatches 82;	Indels 109;	Gaps 11;

```

QY      104 EDTQOENGSMESLITWNID-----GLDLNLSERARGVCSYLALYSPPVIFLOEV-- 153
      : ||| : |::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3 DQTOQFS---FKVLTINIHKGFTAFNRKFIPELRDAVRTV-----SADIVCLOEVMG 52

QY      154 ---IPPYVSYLKKRSSNYEITG-----HEEGYFTAIMLKSRVCLKSQE-- 195
      : | : |::| | : | | : | | : | | : | | : | | : | | : | | : |
Db      53 AHEVHPLHENWPDTSHYEFLADTMSDSFAYGRNAVYDEGHGNNAVL--SRYPDIEHYENR 110

QY      196 -----IIPFSTKMMRNLLCVHVNVSQNELCLMTSHLESTRGHAERMN 239
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      111 DVSVDGAEKRGVLYCRIVP-PMTGKAIHVMCVHLGL-----REAHROA 152

QY      240 QLKMWLKKMQAPESATVIFAGDTNLRDREVTRCGGLPNNIVDVWEFLGPKRHCOYTWD 299
      || | : : | | : | : | | | | | | | | | | | | | | | | | | | | | |
Db      153 QLAMLAENVNELPDGEPVLVAGDEN-----DWRQ 181

QY      300 QMNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKL 342
      : | | : | | : | | | | | | | | | | | | | | | | | | | | | | |
Db      182 KANHPLKVOAG---LDEIFTTRAHGKPARTFPVOQFPLRLDRI 220

```

RESULT 10
F64815
ybhP protein - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C/Accession: F64815
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: F64815
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-253 <BLAT>
A/Cross-references: GB:AE000181; GB:U00096; NID:g1786998; PIDN:AACT3877.1; PID:g17870
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: ybhP

```

Query Match          5.5%; Score 105.5; DB 2; Length 253;
Best Local Similarity 20.8%; Pred. No. 0.47;
Matches 59; Conservative 33; Mismatches 82; Indels 109; Gaps 11;

QY 104 EDTQÖNGSMFSLITWNID-----GLDLNLSEKRGVCSYLALYSPDVIQLÖEV-- 153
      :|||: |::||| | | | | | | | | | | | | | | | | | | | | | | |
Db 3 DQTQÖFS--FKVLTINIHKGFTAFNRFILPELRDAVRTY-----SADIYQLÖEVMG 52

QY 154 ---IPPYVSYLKKRSSNVEIITG-----HEGQYTAIMLKRSRYALKSQE-- 195
      : | : :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 53 AHEVHPLHVENWPDTSHYEFLADTMSWSDFAYGKNAVYPEGHGNAYL--SRPEIEHYENR 110

QY 196 -----IIPFSTKMMRNLLCVHNVNMSGNELCLMTSHLESTRGHAERMN 239
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 111 DVSVDGAEKRGVLYCRIVP-PMTGKAIHVMCVHLGL-----REAHKQA 152

QY 240 QLKMYLKKMQEAPESATVIFAGDTNLRDREYTRCGGLEPNNIIVDWVEFLGKPKHCQYTWD 299
      |||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 153 QLAMLAEWYNELPDGEPVLVAGDEN-----DWRQ 181

QY 300 QMNSNLGITAAACKLRFDRIFERRAAAEHGHIIPRSLDLGLEKL 342
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 182 KANHPLIKYVAG---LDEIFTRAGRPARTFPVQFPLLRDRI 220

```

RESULT 11
T46340

hypothetical protein DKEZp434B0814.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46340
R;Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: 223037
A;Accession: T46340
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-673 <AAA>
A;Cross-references: EMBL:AL137268
A;Experimental source: adult testis; clone DKEZp434B0814
C;Genetics:
A;Note: DKEZp434B0814.1

Query Match 5.5%; Score 105.5; DB 2; Length 673;
Best Local Similarity 20.4%; Pred. No. 1.9;
Matches 78; Conservative 62; Mismatches 135; Indels 107; Gaps 20;

QY 7 LEGGREA-EEGEPEVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNSYF--- 62
Db 145 VEGSMWAALPMQSEPP-----YADC-AALPVGALATEQWEEDPAVLAWSTAP 190
QY 63 EP-PVEESAL-----ERRPETISPKTYVDLTNETTDTSTSKISPSEDTQENGSMF 114
Db 191 EPPVQEEASIWPEEGLGQLPPAVEIP--YHEILWREWEDSTQPDAGLKAGDGPQFQF 248
QY 115 SLITWNIDGLDIN-----NLSEARAGVCSYLALYSPDVIFLOEVIIPPYY- 158
Db 249 TLMSTYNILAQDLMQSSELYLHCHPDLIMWNYRFVNLMEQEFQHWDPDILCIQEVQEDHYW 308
QY 159 -----SYLKKRSSNYEITG-HEEGYFTAIMLKSRVKLSQELIPF--PST 202
Db 309 EOLEPSLRMGFTCFYKRR-----TGCKTGD--CAVCYKPTFRRLCASPVEYFRPGL 359
QY 203 KMMRN-----LLCVH----NVSANELCLMTSHLESTRGHAERMNQLKMKMQE 250
Db 360 ELLNRDNLVGLLQPLVPEGLGQVAVPLCVANTHILNPRRGDVKLAQMALILAEVDK 419
QY 251 AP-----ESATVIFAGDTN-----LRDREVTRCGGLPNNIVDWEFGLKPKHCQY 295
Db 420 VARLSDGSHCPITLLCGDLNSVPDSPLYNFIIRDGEL-QYHGMP-----AWKVSQGEDEFHQ 473
QY 296 TWDTOQN-----SNLGTACK 312
Db 474 LYORKLQAPLWPSLSGITDCQ 495

RESULT 12
A84888
hypothetical protein At2g45230 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cross)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84888
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84888
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1374 <STO>
A;Cross-references: GB:AE002093; NID:g2583130; PIDN:AAB82639.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g45230
A;Map position: 2

Query Match 5.5%; Score 105; DB 2; Length 1374;
Best Local Similarity 24.4%; Pred. No. 6;
Matches 66; Conservative 36; Mismatches 98; Indels 70; Gaps 15;

QY 116 LITWNIDGL----DLNLSERAGVCSYLALYSPDVIFLOEVIIPPYYSYLKKRSSNYEIT 171
Db 3 ILSWNCQGVGNTPTVRHRE-IRG-----LYFEVIFLCET-----KKRRNYLENV 47
QY 172 TGERGYF-----TAIMLKSRVKLSQELIPFPSTKMMRNLLCVHVNVSQN 218
Db 48 VGH-LGFEDLHTEVPIKSGGLALMWKDS-VQIKVLQ-----SDKRLDALLIWQDEFEY 100
QY 219 ELCLMTSHLESTRGHAERMNQLKMKQEAPEASATVIFAGDTNLRDREVTRCGGLPN 278
Db 101 LTCIYGEPPVQAERGELMERLRLGL-----SRSGPWLMTGDENELVDPSEKIGCPAR 152
QY 279 NIVDWEFGLKPKHC-----QYTWDTQMSNLGITACKLRPDRIFFRAAEFGH 328
Db 153 KESSCLEFRQMLNSCGLWEVNHSGYQFSWYGNRDEL-----VQCRLDRTVANQAMME-- 205
QY 329 IIPRSIDLGLLEKLDGCRPPSDHWGLLNL 358
Db 206 LFPQA-KATYLOKI-C-----SDHSPLINNLT 229

RESULT 13
AG2333
hypothetical protein alr4222 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AG2333
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2333
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-281 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA875921.1; PID:g17133357; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4222

Query Match 5.5%; Score 104.5; DB 2; Length 281;
Best Local Similarity 23.1%; Pred. No. 0.66;
Matches 72; Conservative 44; Mismatches 99; Indels 97; Gaps 18;

QY 88 TNEETDSTTSKISPSEDTQOE-NGSMFSLITWNIDGLDNLNLSF-RARGVCSYLALYSP 145
Db 27 TQELTIDS-----SQGLQTELNSNSIKVLNWNLT---AKNFEKFWQDFKILRLYQF 76
QY 146 DVIFLOEVIIPPYYSYLKKRSSNYEITG-----HEEGYFTAIMLKSRVK 190
Db 77 DIIFLOEV-----RMGVNVEQIMGFTNMSWAYAPNFIDAHHQTY--SGILTAAKIS 125
QY 191 LKSQELIPF---PSTKMMRNLLCVHVNVSANELCLMT--SHLESTRGHAERMNQLKMY 244
Db 126 AIAKKVIVTKHHEPVFKTPKVSILITEYSLSHQOQTLTINSHLIN----FVDLDKFKAQ 180
QY 245 LKKMQEA--PESATVIFAGDTN-----LRDREVTRCG-----GLPNNIVDWEFGLK 289
Db 181 LHELELAISTHRGSILFAGDFNTWSRKRAVLLEKTVIRGLKEAHFPEENKKIKRFLLS 240
QY 290 PKHCQYTWDTQMSNLGITACKLRPDRIFFRAAAEEGHIIPRSLDLGLEKLDGGRFPS 349
Db 241 PP-----LDYIFRGLSEK-----PATAKV--LDEI-CS---S 267
QY 350 DHWGLLCNLDI 361
Db 268 DHKPLLAFFTYI 279

RESULT 14

AD1115
hypothetical protein lmo0323 [imported] - *Listeria monocytogenes* (strain EGD-e)
C:Species: *Listeria monocytogenes*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1115
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.:Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00650.1; PID:g16409687; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0323

Query Match	5.4%;	Score 103.5;	DB 2;	Length 256;
Best Local Similarity	25.4%;	Pred. No. 0.7;		
Matches	45;	Conservative	33;	Mismatches 66;
			Indels	33;
			Caps	9;

```
QY      113 MESLITWNIDGLDNLNLSERARG-----VCSYLALYSPDVIFLOEVIPYPYSYLEKRS   165
          ||| : ||| : | | | : : | : | :
Db      1 MFSVTTFNI--REDDTSEKKSKSWELRKLTLSLLDKYQWDFMGVEEPLLPQMRDMKEMQ   57
```

```
QY 166 S-NYEII---TGHEGYFTAIMLKSRVKL-----KSQETIPEPSTKMMRNLLCY- 211
      : :      | : | : | : | :      : :      : :      : :
Db 58 DWDYFGVGGRDDGFEKGEFTAVFYNSTRTFLIQEGHFWLSETPPDVYSIHSTAMFPR-ICW 116
```

```

QY      212  --HVNVSQNELCMTSHLESTRGHAAE--RMNQLKVMKMQEAPESATVIFAGDPTN 264
      : | : : : : | : | : : : : | : : | : |
Db      117  GKPADSDGKQFYIFNTHLD---HISEEARLFASQLLKAATIENSPIIILGDN 169

```

RESULT 15

T24066
exonuclease III homolog - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T24066; T42391

R:Kershaw, J.
submitted to the EMBL Data Library, October 1996

A; Reference number: Z19837

A;Accession: T24066
A;Status: preliminary; translated from GB/EMBL/DBJ

A;molecule type: DNA

A;Residues: 1-288 <WIL>
A;Cross-references: EMBL:Z81108; P1DN:CAB03235.1; GSPDB:GN00019; CESP:R09B3.1

A;Experimental source: clone R09B3
R;Tremblay, S.; Masson, J.Y.; Ramotar, D.

submitted to the EMBL Data Library, November 1997

A;Reference number: Z22150
A;Accession: T42391

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA

A;Residues: 1-288 <TRE>

A;Cross-references: EMBL:AF034258; PIDN:AAC82328.1
C;Genetics:

A;Gene: R09B3.1
A;Map position:

A; Introns: 83/3; 192/3; 264/3
C; Superfamily: exodeoxyribonuclease III

Query Match	5.48;	Score 103;	DB 2;	Length 288;
Best•Local Similarity	21.8%;	Pred. No. 0.91;		
Matches 47;	Conservative 33;	Mismatches 90;	Indels 46;	Gaps 8;

QY	72	ERRPETISEPKTYVDLTNEETDSTSTKISPSSEDTOQENGSMFSLITWNIDGLDLNNLSE	131
	:	: :	:
Dd	3	KRKAEEAPAPK-----LASIFTKKVPAE--EDNNQSKMFCVCMNVAGL-----	44
QY	132	RA---RGVC SYLA LYS PDVI FLQEV---IPPYSYLKRRSSNVEIITGHEEGYFTAILML	184
	:	: :	:
Dd	45	RACVKKSDFKEVLAEPPDLVFLGETCKCEWPEMEETFKNYTKTLVVSTEKNGGYAGVGL	104
QY	185	KKSRVCLKSQEIIPFPSTKMNRLLC-----VHVNVSGNELCLMTSHLESTRGHA	234
	:	: :	:
Dd	105	LSKCAPMKVHKIGIDPEFDTAGRLIIAEFSKFYFIGAYVPNSGAKLV-----NL E-----	154
QY	235	AERMNQ LKVM LKKMQEAPE SAT VIFAGDTNL RDREV	270
	:	: ~ : :	:
Dd	155	-KRGRWEKLLETEKMKEMDEKKPV IYGGDLNVAHNEI	189

Search completed: November 14, 2002, 10:28:15
Job time : 16.5673 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 ; Search time 29.293 seconds
(without alignments)
1646.697 Million cell updates/sec

Title: US-09-697-863A-2
Perfect score: 1902
Sequence: 1 MELGSCLEGGRAAEFEEGEP.....DCGRFPSDHWGLCNLDIIL 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1902	100.0	362	21	AAV56019	Human CD40 recepto
2	1902	100.0	362	22	AAM39841	Human polypeptide
3	1902	100.0	372	22	AAM41627	Human polypeptide
4	1902	100.0	392	20	AAV03182	Topoisomerase II b
5	1894	99.6	369	21	AAB53403	Human colon cancer
6	1893	99.5	362	22	AAB93674	Human protein sequ
7	1609	84.6	306	20	AAV03181	Topoisomerase II b
8	1258	66.1	370	21	AAV56020	Mouse CD40 recepto
9	1199	63.0	263	22	ABG22067	Novel human diagno
10	1150.5	60.5	311	22	ABG22068	Novel human diagno

11	837	44.0	161	20	AAV11868	Human 5' EST seque
12	376	19.8	76	20	AAV12175	Human 5' EST seque
13	231	12.1	404	21	AAV26231	Arabidopsis thalia
14	231	12.1	404	21	AAV42551	Arabidopsis thalia
15	231	12.1	426	21	AAV26230	Arabidopsis thalia
16	231	12.1	426	21	AAV42550	Arabidopsis thalia
17	231	12.1	437	21	AAV26229	Arabidopsis thalia
18	231	12.1	441	21	AAV42549	Arabidopsis thalia
19	169	8.9	63	22	ABG05718	Novel human diagno
20	127.5	6.7	449	22	ABV60974	Drosophila melanog
21	111	5.8	536	22	AAV92307	C glutamicum prote
22	109.5	5.8	1052	22	ABG23154	Novel human diagno
23	106	5.6	1107	21	AAV29693	Mouse apoptosis-re
24	106	5.6	1962	21	AAV29694	Mouse F1ASH protei
25	105.5	5.5	577	22	ABG11645	Novel human diagno
26	105.5	5.5	670	22	AAV24059	Human EST encoded
27	105.5	5.5	670	22	AAV03833	Human gene 16 enco
28	105.5	5.5	670	23	ABG64555	Human albumin fusi
29	105.5	5.5	945	22	ABG06885	Novel human diagno
30	105.5	5.5	1284	22	ABG06053	Novel human diagno
31	105.5	5.5	1284	22	ABG09636	Novel human diagno
32	105.5	5.5	1284	22	ABG10795	Novel human diagno
33	105.5	5.5	1284	22	ABG14889	Novel human diagno
34	105.5	5.5	1703	22	ABG12526	Novel human diagno
35	104.5	5.5	710	22	ABG20795	Novel human diagno
36	104.5	5.5	1557	22	ABG02512	Novel human diagno
37	104.5	5.5	1681	22	ABG01407	Novel human diagno
38	104.5	5.5	1709	22	ABG02341	Novel human diagno
39	104.5	5.5	1709	22	ABG03052	Novel human diagno
40	104.5	5.5	1709	22	ABG03979	Novel human diagno
41	104.5	5.5	1709	22	ABG06055	Novel human diagno
42	104.5	5.5	1709	22	ABG06602	Novel human diagno
43	104.5	5.5	1709	22	ABG08635	Novel human diagno
44	104.5	5.5	1709	22	ABG09638	Novel human diagno
45	104.5	5.5	1709	22	ABG10797	Novel human diagno

ALIGNMENTS

+ RESULT 1			
AAV56019	standard; Protein; 362 AA.		
ID	AAV56019;		
AC	AAV56019;		
DT	15-MAR-2000 (first entry)		
XX			
DE	Human CD40 receptor associated protein.		
XX			
KW	Antiarteriosclerotic; antiarthritic; neuroprotective; dermatological;		Human 5' EST seque
KW	immunosuppressive; antiinflammatory; immunosuppressive; antiallergic;		Human 5' EST seque
KW	human; CD40 receptor associated protein; CRAP; cytoplasmic domain;		Arabidopsis thalia
KW	tumour necrosis factor; TNF; receptor; superfamily; CD30; homology;		Arabidopsis thalia
KW	TNF receptor associated factor; TRAF; modulator; signalling pathway;		Arabidopsis thalia
KW	diagnosis; NF-kappaB; Jun; kinase; atherosclerosis; multiple sclerosis;		Arabidopsis thalia
KW	arthritis; systemic lupus erythematosus; graft rejection; allergy;		Arabidopsis thalia
KW	graft versus host disease; autoimmune disease.		Novel human diagno
XX			
OS	Homo sapiens.		
XX			
PN	W09955859-A2.		
XX			
PD	04-NOV-1999.		
XX			
PF	28-APR-1999; 99WO-EP03025.		
XX			
PR	29-APR-1998; 98EP-0201392.		
XX			
PA	(VLAAs) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.		
XX			
PI	Pyre SMC, Remacle JEFFG, Huylebroeck DFE;		
XX			

DR WPI: 2000-062029/05.
DR N-PSDB; AAZ47119.
XX
PT Novel proteins used to treat inflammatory diseases, NF-kappaB related
PT diseases and for improvement of anti-tumor treatments -
XX
PS Claim 2; Page 39-41; 48pp; English.
XX
CC This sequence represents the human CD40 receptor associated protein
CC (CRAP). CRAP is a functional protein capable of interacting with the
CC cytoplasmic domain of CD40 and/or other receptors of the tumour necrosis
CC factor (TNF) receptor superfamily such as CD30 and TNF receptor I, where
CC the protein has no homology to TNF receptor associated factor (TRAF)
CC proteins. The CD40 binding proteins can be used as modulators of the
CC CD40 signalling pathway, especially to diagnose and treat TRAF-related,
CC CD40-related, NF-kappaB related and/or Jun (kinase)-related diseases,
CC and for the improvement of anti-tumour diseases. Diseases which may be
CC treated include atherosclerosis, arthritis, multiple sclerosis, systemic
CC lupus erythematosus, graft rejection, graft versus host disease, allergy,
CC and autoimmune disease. The proteins can be used to sensitize tumour
CC cells to anti-tumour treatments and to screen for compounds which
CC interfere with the interaction of the proteins with other protein
CC components of the TRAF, CD40 or NF-kappaB related pathway.
XX
SQ Sequence 362 AA;

Query Match 100.0%; Score 1902; DB 21; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.2e-188;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAEEEGEPEVKRRLLCVEFASVASCDAVAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGREAEEEGEPEVKRRLLCVEFASVASCDAVAQCFLAENDWEMERALNS 60
QY 61 YFEPPEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 120
Db 61 YFEPPEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 120
QY 121 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPIYSSYLKKRSSNYEITTGHEGYFT 180
Db 121 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPIYSSYLKKRSSNYEITTGHEGYFT 180
QY 181 AIMLKSRVKLKSQEIIPFPSTKMRNLLCVHNVSGNELCLMTSHLESTRGHAAERMNQ 240
Db 181 AIMLKSRVKLKSQEIIPFPSTKMRNLLCVHNVSGNELCLMTSHLESTRGHAAERMNQ 240
QY 241 LKMYLKKMQEAPESATVTFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQ 300
Db 241 LKMYLKKMQEAPESATVTFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQ 300
QY 301 MNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKLDCCGRFPSDHWGLLCNLDI 360
Db 301 MNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKLDCCGRFPSDHWGLLCNLDI 360
QY 361 IL 362
Db 361 IL 362

^RESULT 2
AAM39841 ID AAM39841 standard; Protein; 362 AA.
XX
AC AAM39841;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2986.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW - chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX WO200153312-A1.
PN
XX
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AA158997.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX
PS Example 4; SEQ ID NO 2986; 10078pp; English.
XX

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Query Match 100.0%; Score 1902; DB 22; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.2e-188;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAEEEGEPEVKRRLLCVEFASVASCDAVAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGREAEEEGEPEVKRRLLCVEFASVASCDAVAQCFLAENDWEMERALNS 60
QY 61 YFEPPEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 120
Db 61 YFEPPEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 120
QY 121 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPIYSSYLKKRSSNYEITTGHEGYFT 180
Db 121 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPIYSSYLKKRSSNYEITTGHEGYFT 180
QY 181 AIMLKSRVKLKSQEIIPFPSTKMRNLLCVHNVSGNELCLMTSHLESTRGHAAERMNQ 240
Db 181 AIMLKSRVKLKSQEIIPFPSTKMRNLLCVHNVSGNELCLMTSHLESTRGHAAERMNQ 240

CC This sequence represents the topoisomerase II binding protein (TopBP) of
CC the invention. The TopBP protein is useful as an anticancer agent. TopBP
CC can be used as the target molecule for anticancer agent.
XX

XX Sequence 392 AA;

Query Match 100.0%; Score 1902; DB 20; Length 392;
Best Local Similarity 100.0%; Pred. No. 3.7e-188;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 60
Db 31 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 90
QY 61 YFEPVVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 120
Db 91 YFEPVVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 150
QY 121 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
Db 151 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 210
QY 181 AIMLKSRVKLKLSQEIIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
Db 211 AIMLKSRVKLKLSQEIIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 270
QY 241 LKMVLKKMQAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQ 300
Db 271 LKMVLKKMQAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQ 330
QY 301 MNSNLGITTAACKLRFDRIFRRAAAEHGHIIPRSLDLGLLEKLDGGRFSPDHWGLLCNLDI 360
Db 331 MNSNLGITTAACKLRFDRIFRRAAAEHGHIIPRSLDLGLLEKLDGGRFSPDHWGLLCNLDI 390
QY 361 IL 362
Db 391 IL 392

*RESULT 5

AAB53403
ID AAB53403 standard; Protein; 369 AA.

XX AC AAB53403;

DT 09-MAR-2001 (first entry)

XX DE Human colon cancer antigen protein sequence SEQ ID NO:943.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.

XX OS Homo sapiens.

XX PN WO200055351-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000MO-US05883.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2000-587534/55.

XX DR N-PSDB; AAC98160.

XX PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX

PS Claim 11; Page 1508-1509; 2104pp; English.

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX SQ Sequence 369 AA;

Query Match 99.6%; Score 1894; DB 21; Length 369;
Best Local Similarity 99.7%; Pred. No. 2.2e-187;
Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 60
Db 8 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 67

QY 61 YFEPVVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 120
Db 68 YFEPVVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQENGSMFSLITWN 127

QY 121 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
Db 128 IDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 187

QY 181 AIMLKSRVKLKLSQEIIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
Db 188 AIMLKSRVKLKLSQEIIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 247

QY 241 LKMVLKKMQAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQ 300
Db 248 LKMVLKKMQAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQ 307

QY 301 MNSNLGITTAACKLRFDRIFRRAAAEHGHIIPRSLDLGLLEKLDGGRFSPDHWGLLCNLDI 360
Db 308 MNSNLGITTAACKLRFDRIFRRAAAEHGHIIPRSLDLGLLEKLDGGRFSPDHWGLLCNLDI 367

QY 361 IL 362
Db 368 IL 369

*RESULT 6

AAB93674

ID AAB93674 standard; Protein; 362 AA.

XX AC AAB93674;

DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:13210.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13210; 2537pp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX

SQ Sequence 362 AA;

Query Match 99.5%; Score 1893; DB 22; Length 362;
Best Local Similarity 99.7%; Pred. No. 2.8e-187;
Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MELGSCLEGREAEEGEPEVKRRRLCYEFASVASCDAAVAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGREAEEGEPEVKRRRLCYEFASVASCDAAVAQCFLAENDWEMERALNS 60
OY 61 YFEPVVEESALERRPETISEPKTYVDLTNEETTDSTTSKISPSEDTQENGSMFSLITWN 120
Db 61 CFEPVVEESALERRPETISEPKTYVDLTNEETTDSTTSKISPSEDTQENGSMFSLITWN 120
OY 121 IDGLDLNLSERARGVSYLALYSPDVIFLQEVIPPYYSYLKKRSSNYEITTGHEGYFT 180
Db 121 IDGLDLNLSERARGVSYLALYSPDVIFLQEVIPPYYSYLKKRSSNYEITTGHEGYFT 180
OY 181 AIMLKRSRVKLKSQETIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
Db 181 AIMLKRSRVKLKSQETIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
OY 241 LKMWLKKMQEAPESATVIFAGDTNLRDREVTRCGGLPNNIVDVWEFLGPKKHQYTWDTQ 300
Db 241 LKMWLKKMQEAPESATVIFAGDTNLRDREVTRCGGLPNNIVDVWEFLGPKKHQYTWDTQ 300

Db 241 LKMWLKKMQEAPESATVIFAGDTNLRDREVTRCGGLPNNIVDVWEFLGPKKHQYTWDTQ 300
OY 301 MNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKLDGCRFPSDHWGLICNLDI 360
Db 301 MNSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKLDGCRFPSDHWGLICNLDI 360
OY 361 IL 362
Db 361 IL 362

RESULT 7
AAV03181
ID AAV03181 standard; protein; 306 AA.
XX
AC AAV03181;
XX
DT 16-JUN-1999 (first entry)
XX
DE Topoisomerase II binding protein.
XX
KW Topoisomerase II binding protein; TopBP; anticancer agent.
XX
OS Homo sapiens.
XX
PN JP11075856-A.
XX
PD 23-MAR-1999.
XX
PE 17-SEP-1997; 97JP-0251544.
XX
PR 17-SEP-1997; 97JP-0251544.
XX
PA (CHUS) CHUGAI PHARM CO LTD.
PA (TSUR/) TSURUO T.
XX
DR WPI; 1999-257704/22.
XX

PT New Topoisomerase II- binding protein - useful as an anticancer
PT agent
XX
PS Claim 1; Page 12-13; 28pp; Japanese.
XX

CC This sequence represents the topoisomerase II binding protein (TopBP) of
CC the invention. The TopBP protein is useful as an anticancer agent. TopBP
CC can be used as the target molecule for anticancer agent.
CC
XX

SQ Sequence 306 AA;

Query Match 84.6%; Score 1609; DB 20; Length 306;
Best Local Similarity 100.0%; Pred. No. 6e-158;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 57 ALNSYFEPVVEESALERRPETISEPKTYVDLTNEETTDSTTSKISPSEDTQENGSMFSL 116
Db 1 ALNSYFEPVVEESALERRPETISEPKTYVDLTNEETTDSTTSKISPSEDTQENGSMFSL 116
OY 117 ITWNIDGLDLNLSERARGVSYLALYSPDVIFLQEVIPPYYSYLKKRSSNYEITTGHEE 176
Db 61 ITWNIDGLDLNLSERARGVSYLALYSPDVIFLQEVIPPYYSYLKKRSSNYEITTGHEE 120
OY 177 GYFTAIMLKRSRVKLKSQETIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAEE 236
Db 121 GYFTAIMLKRSRVKLKSQETIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAEE 180
OY 237 RMNQLKMWLKKMQEAPESATVIFAGDTNLRDREVTRCGGLPNNIVDVWEFLGPKKHQY 296
Db 181 RMNQLKMWLKKMQEAPESATVIFAGDTNLRDREVTRCGGLPNNIVDVWEFLGPKKHQY 240
OY 297 WDTQMSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKLDGCRFPSDHWGLIC 356
Db 241 WDTQMSNLGITACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKLDGCRFPSDHWGLIC 300

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 263 AA;

Query Match 63.0%; Score 1199; DB 22; Length 263;
Best Local Similarity 98.7%; Pred. No. 1.6e-115;
Matches 226; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 133 ARGVCSYLALYSPDVIFLQEVIPYYSYLKKRSSNYEIIITGHEGYFTAIMLKSRVKLK 192
Db 1 ARGVCSYLALYSPDVIFLQEVIPYYSYLKKRSSNYEIIITGHEGYFTAIMLKSRVKLT 60

OY 193 SQEITPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQLKMKVKKMQEAP 252
Db 61 SQEITPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQLKMKVKKMQEAP 120

OY 253 ESATVTFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQMSNLGITAACK 312
Db 121 ESATVTFAGDTNLRDREVTTCGGLPNNIVDVWEFLGKPKHCQYTWDTQMSNLGITAACK 180

OY 313 LRFDRIFFRAAAEEGHIIIPRSLDLGLEKLDGGRFPSDHWGLLCNLDII 361
Db 181 LRFDRIFFRAAAEEGHIIIPRSLDLGLEKLDGGRFPSDHWGLLCNLDII 229

4 RESULT 10
ABG22068
ID ABG22068 standard; Protein; 311 AA.

AC ABG22068;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #22059.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS86255.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID No 52427; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 311 AA;

Query Match 60.5%; Score 1150.5; DB 22; Length 311;
Best Local Similarity 66.3%; Pred. No. 2.2e-110;
Matches 242; Conservative 13; Mismatches 29; Indels 81; Gaps 6;

OY 1 MELGSCLEGGREAAEEGEPEVKKRRLLCVFASVASCAAAVQCFLAENDWEMERALNS 60
Db 25 MELGSCLYGFREAAEEGEPEVKKRRLLCVFASVASCAAAVQCFLAENDWEMERALNS 84

OY 61 YFEPVPEESALERREPTISEPKTYVDLTNEETTDSTTSKISPSEDTQOENGSMFSLITWN 120
Db 85 YFEPPEESALERREPTISEPKTYVDLTNEETTDSTTSKISPSEDTQOENGSMFSLITWE 144

OY 121 -IDGLDL-NNLSEARAGVCSYLALYSPDVIFLQEVIPYYSYLKKRSSNYEIIITGHEGY 178
Db 145 YLMGLDLKQSVQRRARAGVCSYLALXVLSLFNSNV-----SYV----- 182

OY 179 FTAIMLKSRVKLKSGQEIIPFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERM 238
Db 183 -----YTERM 187

OY 239 NQLKMKVKKMQEAPESATVIFAGDTNLRDREVT-RCGGLPNNIVDVWEFLGKPKHCQYTW 297
Db 188 NQLKMKVKKMQEAPESATVIFAGDTNLRDREVS DKVSLMFEDRLLMLRLIXP-FCQYTW 246

OY 298 DTQMSNLGITAACKLRFDRIFFRAAAEEGHIIIPRSLDLGLEKLDGGRFPSDHWGLLCN 357
Db 247 DTQMSNLGITAACKLRFDRIFFRAAAEEGHIIIPRSLDLGLEKLDGGRFPSDHWGLLCN 306

OY 358 LDITL 362
Db 307 LDITL 311

RESULT 11
AAY11868
ID AAY11868 standard; Protein; 161 AA.

AC AAY11868;

DT 18-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID No: 468.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.
OS
XX
XX
PN W09906550-A2.

PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01232.
XX
PR 01-AUG-1997; 97US-0905144.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153780/13.
DR N-PSDB; AAX40590.
XX
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Claim 34; Page 595; 675pp; English.

CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11993 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 161 AA;

Query Match 44.0%; Score 837; DB 20; Length 161;
Best Local Similarity 100.0%; Pred. No. 2.4e-78;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGGRAAEEEGEPEVKRRLLCVEFASVASCDAVAQAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGGRAAEEEGEPEVKRRLLCVEFASVASCDAVAQAQCFLAENDWEMERALNS 60
QY 61 YFEPVVEESALERPETISEPKTYVDLTNETTDTSTKISPSSEDTQOENGSMFSLITWN 120
Db 61 YFEPVVEESALERPETISEPKTYVDLTNETTDTSTKISPSSEDTQOENGSMFSLITWN 120
QY 121 IDGLDLNNLSERARGVCSYLALYSPDVIFLQEVIPPYSYL 161
Db 121 IDGLDLNNLSERARGVCSYLALYSPDVIFLQEVIPPYSYL 161

RESULT 12
AAY12175
ID AAY12175 standard; Protein; 76 AA.
XX
AC AAY12175;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human 5' EST secreted protein SEQ ID NO: 488.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.

XX
OS Homo sapiens.
XX
PN WO9906554-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01238.
XX
PR 01-AUG-1997; 97US-0905134.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153784/13.
DR N-PSDB; AAX41008.
XX

PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
PS Claim 34; Page 567; 622pp; English.

CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY01602 and
CC AAY11994 to AAY12260, respectively. The proteins given represent the
CC signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.

XX Sequence 76 AA;

Query Match 19.8%; Score 376; DB 20; Length 76;
Best Local Similarity 96.0%; Pred. No. 4.6e-31;
Matches 72; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELGSCLEGGRAAEEEGEPEVKRRLLCVEFASVASCDAVAQAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGGRAAEEEGEPEVKRRLLCXEFXSVASCDAVAQAQCFLAXNDWEMERALNS 60
QY 61 YFEPVVEESALERRP 75
Db 61 YFEPVVEESALERRP 75

RESULT 13
AAG26231
ID AAG26231 standard; Protein; 404 AA.
XX
AC AAG26231;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 30610.

XX Arabidopsis thaliana protein fragment SEQ ID NO: 30610.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.

XX
PN . EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140981.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142055.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.

PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158032
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

Query Match	12.1%;	Score 231;	DB 21;	Length 404;
Best Local Similarity	30.2%;	Pred. No. 7.9e-15;		
Matches 84; Conservative	39;	Mismatches 109;	Indels 46;	Gaps 11

```

QY      113 MSLITWNIDGLDNLNLSERAGVCSYIALYSPDVIQLQEVIPPIYSYLKRR-----SS 166
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      138 ILSYNWVFREDLELN---LRMRAIGHLIQLHSPHLICFOEVTPETLIDIFRKSNNWKKAYSC 194

QY      167 NYEIIITGHEEGYFTAIMLKSSRVKLKSQELIIPFPSTKMMRNLCYHVNVSNGNE-ICIMTS 225
      : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      195 SVSVDVAVSRGYCYCMLL--SKLGVKSFSSKSGNSIMGRELSIAEVEVPGRKPLVFATS 251

QY      226 HLES-----TRGHAERMNOLKMWLKKQEAPEASATVIFAGDTNLRDREYTRCGG-- 275
      |||| : : | : | : | : : : : | |||| | : | : | : | : | : | : |
Db      252 HLESPCGPRPKWDQMFSSREREVEQAKAEIFILR--PNANVIFGGDMNWCD---KLDGKF 304

QY      276 -LPNNIVDVWEFLGKPKHCQYTWDTOMNSNIGITAAACKLRFDRIFER----- 321
      ||: |||| | | : : ||: | | : | ||| |
Db      305 PLPDKWDVDWEVL-KPGDLGFTYDTKANPMLSGNRALQKRDLRIICRLDDYKLGIEIENWG 363

QY      322 AAEEGHIIPRSLDLIG-LEKLDGGRFPSPDHWGILCNL 358
      | | : : | : ||: ||||: ||| |

```


DQ 364 KEAIPGLSYVKEKKVRGDIKKLELPVLPSDHGLLVTL 401

RESULT 15

AAG26230

ID AAG26230 standard; Protein; 426 AA.

XX

AC AAG26230;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 30609.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139763.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 28-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 231; DB 21; Length 426;
Best Local Similarity 30.2%; Pred. No. 8.7e-15;
Matches 84; Conservative 39; Mismatches 109; Indels 46; Gaps 11;

OY 113 MFSLITWNIDGLDNLNLSERAGVCSYLALYSPDVIFLQEVIPPYSYLKKR-----SS 166
Db 160 ILSYNVWFERDELELN--LRMRALGHLLIQLHSPHLICFQEVTPREIYDIFRKSMMWKAYSC 216
OY 167 NYELITGHEEGYFTAIMLKKSRYVILKSQEIIPFPSTKMMRNLLCVHVNVSQNE-LCLMTS 225
Db 217 SVSVDVAVSRGYCMLL---SKLGVKSFSSSKSGFNSTMGRELISIAEVEVPGRKPLVFATS 273
OY 226 HLES-----TRGHAARMNDLKMVLKKMQEAPESATVIFAGDTNLRDRREVTROCG-- 275

Db 274 HLESPCPGPPKWDQMFSRREVEQAKEAIEILR---PNANVIFGGMNWC-----KLDGKF 326
OY 276 -LPNNIVDVWEFLGKPKHCQYTWDTQMSNLGITACKLRFDRIFFR----- 321
Db 327 PLPDKWVDWEYL-KPGDLGFTYDTKANPMLSGNRALQKRLDRILCRLLDDYKLGIEWVG 385
OY 322 AAEEGHIIPRSLDLG-LEKLDGGRFPPSDHWGLICNL 358
Db 386 KEAIPGLSYVKEKKVRGDIKKLELPVLPSPSDHFGLLVTL 423

Search completed: November 14, 2002, 10:27:24
Job time : 31.293 secs

Gencore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 26.1584 Seconds
(without alignments)
2914.450 Million cell updates/sec

Title: US-09-697-863A-4
Perfect score: 1968
Sequence: 1 MASGSSSDAAEPAGPAGRAA.....DCGRFSDHWGLCTLNVVL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1968	100.0	370	11 Q9JUX7	Q9jix7 mus musculu
2	1258	63.9	362	4 Q9NYI9	Q9nyy9 homo sapien
3	1258	63.9	362	4 Q95551	Q95551 homo sapien
4	1249	63.5	362	4 Q9NNU5	Q9nuk5 homo sapien
5	849	43.1	157	11 Q9D7N0	Q9d7n0 mus musculu
6	432.5	22.0	362	5 Q9XWG3	Q9xwg3 caenorhabdi
7	275	14.0	441	10 Q9SA95	Q9sa95 arabidopsis
8	125.5	6.4	449	5 Q9VGS4	Q9vgs4 drosophila
9	118.5	6.0	253	16 Q9S2L9	Q9s2l9 streptomyce
10	111.5	5.7	252	16 Q92X70	Q92x70 rhizobium m
11	102	5.2	265	16 Q8UEP1	Q8uep1 agrobacteri
12	102	5.2	335	2 Q9RLV9	Q9rlv9 listeria iv
13	101.5	5.2	405	5 Q15921	Q15921 trypanosoma
14	101	5.1	1114	5 Q27811	Q27811 tripeustes
15	100.5	5.1	502	3 Q9Y7M8	Q9y7m8 schizosacch
16	100.5	5.1	648	2 Q8RQ53	Q8rq53 leptospira

17	100	5.1	329	2 Q68855	Q68855 staphylococ
18	99.5	5.1	678	10 Q9C8J8	Q9c8j8 arabidopsis
19	99	5.0	327	5 Q9U2B5	Q9u2b5 caenorhabdi
20	98.5	5.0	765	16 Q8ZQ02	Q8zq2 salmoneilla
21	97.5	5.0	644	4 Q9NRR6	Q9nrr6 homo sapien
22	97	4.9	447	10 Q9FFQ0	Q9ffq0 arabidopsis
23	97	4.9	3571	10 Q9SL27	Q9sl27 arabidopsis
24	97	4.9	3574	10 Q9AUB4	Q9aub4 arabidopsis
25	96.5	4.9	840	16 P72843	P72843 synechocyst
26	96	4.9	626	2 P77988	P77988 thermoaer
27	96	4.9	633	10 Q9LL45	Q9ll45 oryza sativ
28	95	4.8	259	16 Q8ZRM4	Q8zrm4 salmoneilla
29	94.5	4.8	549	5 Q60963	Q60963 leishmania
30	94.5	4.8	765	2 Q9XCV2	Q9xcv2 salmoneilla
31	94	4.8	1240	3 Q9P6U5	Q9p6u5 neurospora
32	94	4.8	2911	5 Q9BLV4	Q9blv4 leishmania
33	94	4.8	3040	5 Q9GNY4	Q9gny4 leishmania
34	93	4.7	270	16 Q9I0T9	Q9i0t9 pseudomonas
35	92.5	4.7	288	5 Q45711	Q45711 caenorhabdi
36	91.5	4.6	313	17 Q978R5	Q978r5 thermoplasm
37	91.5	4.6	792	5 Q9NMV3	Q9nmv3 leishmania
38	91.5	4.6	825	11 Q91ZF8	Q91zf8 mus musculu
39	91	4.6	4024	4 Q8WXX0	Q8wxx0 homo sapien
40	90.5	4.6	612	10 Q9LGT1	Q9lgt1 oryza sativ
41	90.5	4.6	617	12 Q83534	Q83534 measles vir
42	90.5	4.6	617	12 Q83537	Q83537 measles vir
43	90	4.6	608	16 Q9A3Q7	Q9a3q7 caulobacter
44	90	4.6	684	12 Q9ENK9	Q9enk9 colorado ti
45	90	4.6	691	17 Q9HPV0	Q9hpv0 halobacteri

ALIGNMENTS

RESULT 1					
Q9JUX7					
ID Q9JUX7	PRELIMINARY;	PRT;	370 AA.		
AC Q9JUX7;					
DT 01-OCT-2000 (TREMBlrel. 15, Created)					
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)					
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)					
DE Putative TRAF and TNF receptor associated protein.					
GN TRAF OR TRAF.					
OS Mus musculus (Mouse).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX NCBI_TaxID=10090;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=20309820; Pubmed=10764746;					
RA Pyre S., Declercq W., Ibrahim A., Michiels C.,					
RA Van Riettschoten A.G.I., Dewulf N., de Boer M., Vandenberghe P.,					
RA Huylebroeck D., Remacle J.E.;					
RT "TRAF, a novel protein that associates with CD40, Tumor Necrosis					
RT Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs),					
RT and that inhibits Nuclear Factor-kappaB activation.";					
RL J. Biol. Chem. 275:18586-18593(2000).					
DR EMBL; AJ251328; CAB92971.1; -.					
DR MGD; MGI:1860486; Ttrap.					
DR InterPro; IPR005135; Exo_endo_phos.					
DR Pfam; PF03372; Exo_endo_phos; 1.					
KW Receptor.					
SQ SEQUENCE 370 AA; 41033 MW; A773A8889DF5BE83 CRC64;					
Query Match	100.0%;	Score 1968;	DB 11;	Length 370;	
Best Local Similarity	100.0%;	Pred. No. 3.2e-168;			
Matches 370;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1 MASGSSSDAAEPAGPAGRAASAPEAQAQAEEDRVKRRRLQCLGFALVGGCPTWPSVLR 60					
Db 1 MASGSSSDAAEPAGPAGRAASAPEAQAQAEEDRVKRRRLQCLGFALVGGCPTWPSVLR 60					
QY 61 NDWOTOKALSAFFELPENDQGWPRQPTSFKSEAYVDLTNEDANDTTILEASPSGTPLD 120					


```
Db      61 NDWQOKALSAYFELPENDQGWPRQPTSEKSEAYVDLTNEDANDTTILEASPSGTPLED 120
QY      121 SSTISFTWNIDGLDGCNLPERARGVCSCLALYSPDVVFLQEVIPPYCAYLKKRAASYT 180
Db      121 SSTISFTWNIDGLDGCNLPERARGVCSCLALYSPDVVFLQEVIPPYCAYLKKRAASYT 180
QY      181 ITGNEGYFTAILLKKGRVKFKSQEIIPEPNTKMMRNLLCVNYSLGNNEFCIMTSHLEST 240
Db      181 ITGNEGYFTAILLKKGRVKFKSQEIIPEPNTKMMRNLLCVNYSLGNNEFCIMTSHLEST 240
QY      241 REHSAERIRQLKTVLGKMQEAPDSTVIFAGDTNLRDQEVIKCGGLPDNVFDAMEFLGKP 300
Db      241 REHSAERIRQLKTVLGKMQEAPDSTVIFAGDTNLRDQEVIKCGGLPDNVFDAMEFLGKP 300
QY      301 KHCQYTWDTKANNNLRIPAAVKHREDRIFFRAEEGHLIPQSLDLVGLKLDGGRFPSDHW 360
Db      301 KHCQYTWDTKANNNLRIPAAVKHREDRIFFRAEEGHLIPQSLDLVGLKLDGGRFPSDHW 360
QY      361 GLLCTLNVL 370
Db      361 GLLCTLNVL 370
```

```
RESULT 2
Q9NYY9 PRELIMINARY; PRT; 362 AA.
AC Q9NYY9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE AD022 protein.
GN AD022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL GLAND;
RA Song H., Gao G., Peng Y., Ren S., Chen Z., Han Z.;
RT "A novel gene expressed in human adrenal gland.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF223469; AAF64144.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40956 MW; D6D607DE521C27E4 CRC64;
```

Query Match 63.9%; Score 1258; DB 4; Length 362;
Best Local Similarity 68.4%; Pred. No. 1.6e-104;
Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;

```
QY      24 EAAQAE-EDRVKRRRLQCLGFALVGGCDPTMVPVSLRENDWQOKALSAYFELPENDQGW 82
Db      12 EAAEEGEPEVKRRLLCVRFASVASCDAVAQCFLAENDWEMERALNSYFEPVVEESAL 71
QY      83 PROPTSEKSEAYVDLTNEDANDTTILEASPS-GTPLEDSSTISFTWNIDGLDGCNLP 141
Db      72 ERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQQENGSMFSLITWNIDGLDNLNLS 131
QY      142 RARGVCSCLALYSPDVVFLQEVIPPYCAYLKKRAASYTIITGNEGYFTAILLKKGRVKF 201
Db      132 RARGVCSYALYSPDVIFLQEVIPPYSYLKKRSSNYEITTGHEGYFTAILMKRSVKL 191
QY      202 KSQEIIPEPNTKMMRNLLCVNYSLGNNEFCIMTSHLESTREHSAERIRQLKTVLGKMQEA 261
Db      192 KSQEIIPEPSTKMMRNLLCVHNVSGNELCIMTSHLESTRGHAAERMNQLKMKMQEA 251
QY      262 PDSTVIFAGDTNLRDQEVIKCGGLPDNVFDAMEFLGPKKHQYTWDTKANNNLRIPAAV 321
Db      252 PESATVIFAGDTNLRDREVIKCGGLPNNIVDVWEFLGPKKHQYTWDTQMNLSGITTAAC 311
QY      322 KHRDRIFFR--AEEGHLIPQSLDLVGLKLDGGRFPSDHWGLLCTLNVL 370
```

```
Db      312 KLREDRIFFRAAEEGHIIIPRSLDLGLEKLDGGRFPSDHWGLLCTLNIDL 362
RESULT 3
ID 095551 PRELIMINARY; PRT; 362 AA.
AC 095551;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE D3J0M3.3 (TRAF and TNF receptor associated protein) (ETSL-associated
DE protein 2) (Hypothetical 40.9 kDa protein).
GN D3J0M3.3 OR TTRAP OR EAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309820; Pubmed=10764746;
RA Pype S., Declercq W., Ibrahim A., Michiels C.,
RA Van Riettschoten A.G.I., Dewulf N., de Boer M., Vandenabeele P.,
RA Huylebroeck D., Remacle J.E.;
RT "TTRAP, a novel protein that associates with CD40, Tumor Necrosis
RT Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs),
RT and that inhibits Nuclear Factor-kappaB activation.";
RL J. Biol. Chem. 275:18586-18593(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Li R., Pei H., Papas T.S.;
RT "EAP2, a Novel Protein Interacting with ETS1.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031775; CAA21141.1; -.
DR EMBL; AJ269473; CAB92966.1; -.
DR EMBL; AF201687; AAG35600.1; -.
DR EMBL; BC017553; AAH17553.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Receptor; Hypothetical protein.
SQ SEQUENCE 362 AA; 40929 MW; 37892E125DB64410 CRC64;
```

Query Match 63.9%; Score 1258; DB 4; Length 362;
Best Local Similarity 68.4%; Pred. No. 1.6e-104;
Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;

```
QY      24 EAAQAE-EDRVKRRRLQCLGFALVGGCDPTMVPVSLRENDWQOKALSAYFELPENDQGW 82
Db      12 EAAEEGEPEVKRRLLCVFEFASVASCDAVAQCFLAENDWEMERALNSYFEPVVEESAL 71
QY      83 PROPTSEKSEAYVDLTNEDANDTTILEASPS-GTPLEDSSTISFTWNIDGLDGCNLP 141
Db      72 ERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQQENGSMFSLITWNIDGLDNLNLS 131
QY      142 RARGVCSCLALYSPDVVFLQEVIPPYCAYLKKRAASYTIITGNEGYFTAILLKKGRVKF 201
Db      132 RARGVCSYALYSPDVIFLQEVIPPYSYLKKRSSNYEITTGHEGYFTAILMKRSVKL 191
QY      202 KSQEIIPEPNTKMMRNLLCVNYSLGNNEFCIMTSHLESTREHSAERIRQLKTVLGKMQEA 261
Db      192 KSQEIIPEPSTKMMRNLLCVHNVSGNELCIMTSHLESTRGHAAERMNQLKMKMQEA 251
QY      262 PDSTVIFAGDTNLRDQEVIKCGGLPDNVFDAMEFLGPKKHQYTWDTKANNNLRIPAAV 321
Db      192 KSQEIIPEPSTKMMRNLLCVHNVSGNELCIMTSHLESTRGHAAERMNQLKMKMQEA 251
```

```
Db      252 PESATVIFAGDTNLRDREYTRCGGLPNNIYDVWEFLGPKKHCQYTWTQTOMNSNLGITAAC 311
QY      322 KHRFDRIEFR--AEEGHILPQSLLDLVLGLEKLDCGRFPSDHWGLLCTLNVL 370
       | ||||| ||| :|||:|||:||||| ||||| ||||| ||| :|||
Db      312 KLRFDRIEFRAAAEEGHIIIPRSLDLIGLEKLDGCRFPSDHWGLLCNDIIL 362

RESULT 4
Q9NUK5 PRELIMINARY; PRT; 362 AA.
ID Q9NUK5 AC Q9NUK5:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CDNA FLJ11306 fis, clone PLACE1010031.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Tagiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuko Y.,
RA Ninomiya K., Iwayanagi T.;
RT "MEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK002168; BAA92119.1; -
DR InterPro; IPRO05135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40869 MW; 8E1702C93BCB62E3 CRC64;

Query Match          63.5%; Score 1249; DB 4; Length 362;
Best Local Similarity 68.1%; Pred. No. 1e-103;
Matches 239; Conservative 43; Mismatches 65; Indels 4; Gaps 3;

QY      24  EAAQA-EDRVKKRRRLQGCLGFALVGCDPTMVP SVLRENDMQTKALSAYFELPENDQGW 82
       |||:| | |||:| |:| | | | | | | | | | | | | | | | | | |
Db     12  EAEEEEGEPEVKKRRLICEVFASVASCDAAVAQCFLAENDWEMERALNSCFEPVEESAL 71

QY      83  PROPTSFKSEAYVDLTNEANDTITILEASPS-GTPLEDSSITSFITWNIDGLDGCNLP 141
       | :| | : | ||||| :| :| | | | | | | | | | | | | | | |
Db     72  ERREPETISEPTYVDLTNETFTDSTSKISPS EDTQQENGSMFS LITWINIDGLDNLNSE 131

QY     142 RARGVCSCIALYSPDVVFLQEVIPPYCAYLKKRASYYTIITGN EEGYFTAILLKGRVKF 201
       ||||| | ||||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     132 RARGVCSYLYSPDIVFLQEVIPPYSYLKKRSSNYEIITGHEGYFTAIMLKKSRYKL 191

QY     202 KSQEIIPEPNTKMNRNLICVNVS LGNEFCMLMTSHLESTRHSAERTIRQLKT VLGKMQEA 261
       ||||| |||:|||||:| | | | | | | | | | | | | | | | | | |
Db     192 KSQEIIPEPSTKMNRNLICVHVNVSGNELCLMTSHLESTRGHAERMNQLKMKVMQEA 251

QY     262 PDSTTVIFAGDTNLRDQEVIKGGGLPDNVFDAMEFLGKPKHCQYTWDTKANNNLRIPAAY 321
       | :| |||||:| | | | | | | | | | | | | | | | | | | | |
Db     252 PESATVIFAGDTNLRDREVTRCGGLPNNIYDVWEFLGPKKHCQYTWTQTOMNSNLGITAAC 311

QY     322 KHRFDRIEFR--AEEGHILPOSLLDI VLGLEKLDCGRFPSDHWGLICTINVL 370
       | ||||| | |||||:| | | | | | | | | | | | | | | | | | |
Db     312 KLRFDRIEFRAAAEEGHIIIPRSLDLIGLEKLDGCRFPSDHWGLLCNDIIL 362

RESULT 5
Q9D7NO PRELIMINARY; PRT; 157 AA.
ID Q9D7NO AC Q9D7NO:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Traf and Tnf receptor associated protein.
```

```

GN      TTRAP.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=TONGUE;
RX      MEDLINE=21085660; PubMed=11217851;
RA      Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA      Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA      Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA      Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA      Kuehl P., Lewis S., Matsuo Y., Nikaïdo I., Pesole G., Quackenbush J.,
RA      Schriml L.M., StaUBLI F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de BonaIdo M.F.,
RA      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA      W ynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA      Hayashizaki Y.;
RT      "Functional annotation of a full-length mouse cDNA collection.";
RL      Nature 409:685-690(2001).
DR      EMBL; AK009089; BAB26063.1; -.
DR      MGD; MGI:1860486; Ttrap.
DR      InterPro: IPR005135; Exo_endo_phos.
DR      Pfam; PF03372; Exo_endo_phos; 1.
SQ      SEQUENCE 157 AA; 17876 MW; 5920850802FAAE84 CRC64;

Query Match          43.1%; Score 849; DB 11; Length 157;
Best Local Similarity 100.0%; Pred. No. 2.4e-68;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0,

QY      214 MMRNLLCVNVS LGNEFCLMTSHLESTREHSAERIRQLKTYLGKMQEAPDSTTVIFAGDT 273
Db      1 MMRNLLCVNVS LGNEFCLMTSHLESTREHSAERIRQLKTYLGKMQEAPDSTTVIFAGDT 60

QY      274 NLRDQEVIKCGGLPDNVFDAMEFLGPKHKCQYTWDTKANNNLRIPAYKHRFDRIFFRAE 333
Db      61 NLRDQEVIKCGGLPDNVFDAMEFLGPKHKCQYTWDTKANNNLRIPAYKHRFDRIFFRAE 120

QY      334 EGHLLPQSLDLVGLKLDGGRFPSPDHWGLLCTLNVVL 370
Db      121 EGHLLPQSLDLVGLKLDGGRFPSPDHWGLLCTLNVVL 157

RESULT 6
Q9XWG3 PRELIMINARY; PRT; 362 AA.
ID      Q9XWG3;
AC      Q9XWG3;
DT      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Y63D3A.4 protein.
OS      Y63D3A.4.
SN      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RL      Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RX      MEDLINE=99069613; PubMed=9851916;
RA      none;

```

```

RT "Genome sequence of the nematode C.elegans: A platform for
RL investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032652; CAA21707.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40893 MW; 7895C47504866E52 CRC64;

Query Match 22.0%; Score 432.5; DB 5; Length 362;
Best Local Similarity 33.5%; Pred. No. 1.9e-30;
Matches 113; Conservative 55; Mismatches 126; Indels 43; Gaps 11;

QY 43 FALVGGCDPTMVPVSLRENDQOTOKALSAFELPNDQGWPROPTSFKSEAYVDLTNED 102
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 52 FAITATDEAFAQSILOVDWDLKALDVF-----YGSEAF-----E 89

QY 103 ANDTILEAS-----PSGTPL---ED--SSTISFITWNIDGLDGCNLPERARGVCSCLALY 153
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 ARSAAVMGASSSMASGAAVMTAEDLKGFEVSVMSWNIDGLDGRSLLRMKAVAHIVKNV 149

QY 154 SPDVYFLQEVIPRYCAYLKKRAASYTITITGNEE-GYFTAILLKKGRVKEKSOEIIPEPNT 212
   :|:|:|:|:|:|:| | | | | | | | | | | | | | | | | | | | | | |
Db 150 NPDIQLQEVVDRDLAPIDKLQSLKYYSNKGCOYTAILVSK-MEDVEKHDVHFQNS 208

QY 213 KMMRNLCVNVSLGNEFCLMTSHLESTREHSAERIRQLKTVLGKMQE---APDSTVIIF 269
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 GMYRTLIILEGSIGGLKVFLLNTHLESTREHRRPQCAQGFQFCMDKVREITIAQNPGALVVF 268

QY 270 AGDTNLRDQEVIKCGGLPDNVFDAMEFLGKPKHCQYTWDTKANNNLRIPAAYKHRDRIE 329
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 GGDNLNRDEVS---VPDGYKDAWEAAGSDNKTFTWDTFKNDNKQGFHGAKMRFDRLY 325

QY 330 FRAEEGHILIPQSLDLVGLEKL-DGGRFPSDHWGLICT 365
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 326 W---SGFLDKVKFTLEGGRQIRSCLCFPSDHWAINAT 359

RESULT 7
Q9SA95 PRELIMINARY; PRT; 441 AA.
ID Q9SA95;
AC Q9SA95;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE F25C20.3 protein.
GN F25C20.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.M., Kremenetskaia I., Luros J., Ngan I., Liu A.,
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007296; AAD30241.1; -.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR001876; Znf_RangDP.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR SMART; SM00547; Znf_RBZ; 2.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.

```

```

DR PROSITE; PS01358; ZF_RANBP2_1; UNKNOWN_1.
SQ SEQUENCE 441 AA; 48995 MW; 6582FAA2D7BD0CA8 CRC64;

Query Match 14.0%; Score 275; DB 10; Length 441;
Best Local Similarity 31.4%; Pred. No. 3.4e-16;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

QY 111 ASPSGTPLEDSTISFITWNIDGLDGCNLPERARGVCSCLALYSPDVYFLQEVIPRYCAY 170
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 ASDSGTPLTCLIKILSYNWFREDLE---LNLRMRAIGHLIQLHSPHLCFQEVTEPEIYDI 219

QY 171 LKKR-----AASYTIITGNEGYTTAILLKKGRVKEKSOEIIPEPNTKMMNLLCVNS 224
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 220 FRKSNMWKAYSCSVSDVAVSRGYCMLLSLKGVKSFSSKS---FGNSIMGRELISAEVE 276

QY 225 LGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGKMQEAPDSTVIIFAGDTNL 275
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 VPGKKPLVFATSHLESPPCPGPKWDQMSRERVEQAKEAIEILR---PNANVIFGDMNW 333

QY 276 RDQEVIKCG---LPDNVFDAMEFLGKPKHCQYTWDTKANNNLRIPAAYKHRDRIEFRA 332
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 334 CD---KLDGKRPPLPKWVDWVEVL-KPGDLGFTYDTKANPMLSGNRALQKRLDRILCRL 388

QY 333 EEGHILIPQSLDLVG-----LEKLDGGRFPSDHWGLICTLN 367
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 389 DDYKL--GGIENWGKEAIPGLSYKEREKVRGDIKLELPVLPDSHFGLLVTLIS 439

RESULT 8
Q9VGS4 PRELIMINARY; PRT; 449 AA.
ID Q9VGS4;
AC Q9VGS4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CG4796 protein.
GN CG4796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferltera S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```


QY 202 -----KSQEIIP-----FPNTKMRNLLCVNVS LGNEFC LMTSHL----- 237
Db 142 QWPIVEKSOHIFQRGGADRLSNKGFAVYKIMKN-----GKPYHIIGTHTPADDSL 192
QY 238 ---ESTREHSAERIRQLKTVLGKMQEAPDSTTVIFAGDTNLRDQEVIKCGGLPDNVFDAM 294
Db 193 ISKDTSTRAIRAEOEQEIQTFIAR-KNIPKDEIIFIGIDL-----VNYG--TDEYHDMF 243
QY 295 EFLGKPKHCQY-----TWDTKANNNLR--IPAAYKHFDRIFERRAEGLIPQS 341
Db 244 KLLNVSSPANFNQOMATWDPNTSMLKESYPKAAPEYLDYIF--VENGHARPHS 295
RESULT 13
ID 015921 PRELIMINARY; PRT; 405 AA.
AC 015921;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE AP-endonuclease.
GN TCAP.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_Taxid=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=99108141; Pubmed=9889272;
RA Perez J., Gallego C., Bernier-Villamor V., Camacho A.,
RA Gonzalez-Pacanowska D., Ruiz-Perez L.M.;
RT "Apurinic/apurimidinic endonuclease genes from the trypanosomatidae
RT leishmania major and Trypanosoma cruzi confer resistance to oxidizing
RT agents in DNA repair-deficient Escherichia coli.";
RL Nucleic Acids Res. 27:771-777(1999).
DR EMBL; U92486; AAD11456.1; -
DR HSSP; P27695; 1E9N.
DR InterPro; IPR004442; ExoDNase_III.
DR InterPro; IPR004808; ExoIII_xth.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR TIGRFAMS; TIGR00195; exoDNase_III; 1.
DR TIGRFAMS; TIGR00633; xth; 1.
KW Endonuclease.
SQ SEQUENCE 405 AA; 45193 MW; 3BAFC1882DB8FCF6 CRC64;
Query Match 5.2%; Score 101.5; DB 5; Length 405;
Best Local Similarity 20.7%; Pred. No. 1.1;
Matches 84; Conservative 37; Mismatches 119; Indels 165; Gaps 22;
QY 86 PPTSFKSEAYVDLTNEDANDTILEASPSGTPLEDSS----- 122
Db 44 PPSRSLNSAGAEATSPNRLAVALTAP--PSDDDTKTEKDIWSQVEPFQRTAAKDF 100
QY 123 ----TISFITWNIDGLDGCNLPERARGVCSCLALYSPDVVFLQEV-----I 164
Db 101 DSKHMLKFTWNVAGLRGL-LRKDDQAIQRLLEECPDALCQETKLNPDPPQNEKLG EV 159
QY 165 PPY-----C-----AYLKRAAS-YTIIT-----GNEEG--- 187
Db 160 PGYREYDVHVCRAKKYSGTRTYIKNTAAAEWKTVTVKGFDTLKSPQDVGHSEGEDEGRVL 219
QY 188 --YFTAILLKGRVKFKSQEIIPF-PNTKMMRNLL---CVNVS LGNEFC LMTSHLESTR 241
Db 220 TTYFGT--QKGSETFALALVNTYIIPNSGMSLERLPRCQKFDL----- 261
QY 242 EHSAERIRQLKTVLGKM-----QEAPDSTTVIFAGDTNLRDQ-----VIKC 283
Db 262 -----RIRQHLCTLGRSCNHDKEEGDAPSLAGFTWAGDINVAE RDRYFAGSYKAMQKC 316
QY 284 GGL-PD-----NVDAWEFLGKPKHCQYT-WDTKANNNLRIPAPAYKHFDRIF 329
Db 317 SGFTPEERASFRETLRVANAVDTFRALYPKAAPVYTFWSARINGRAR--GI GWRLDYFV 373

QY 330 FRAEGLHLPQSLDLVLEKLDGGRFP-----SDHWGLLCTLNVL 370
Db 374 VSA-----ALARHVVDCTMPHYMGSDH-----CPLQMWL 403
RESULT 14
ID 027811 PRELIMINARY; PRT; 1114 AA.
AC 027811;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Dynein heavy chain isolate 7A (EC 3.6.1.3) (Fragment).
GN DYH7A.
OS tripeustes gratilla (Hawaiian sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OC Tripeustes.
OX NCBI_Taxid=7673;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=94243035; Pubmed=8186465;
RA Gibbons B.H., Asai D.J., Tang W.J., Hays T.S., Gibbons I.R.;
RT "Phylogeny and expression of axonemal and cytoplasmic dynein genes in
RT sea urchins.";
RL Mol. Biol. Cell 5:57-70(1994).
DR EMBL; U03978; AAA63592.1; -
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hydrolase.
FT NON_TER 1 1
FT NON_TER 1114 1114
SQ SEQUENCE 1114 AA; 125999 MW; BCF09927EF5688BC CRC64;
Query Match 5.1%; Score 101; DB 5; Length 1114;
Best Local Similarity 21.4%; Pred. No. 5.4;
Matches 53; Conservative 36; Mismatches 87; Indels 72; Gaps 11;
QY 12 PAGPAGRAASAPBEAAQAEEDRVKRRRLQCLGFALVGGCDPTWVPSVLRNDWQTKALSA 71
Db 39 PEGPAGTGKT-----ETTKDLAKAVAKQCVFENCSDGLD-----YIALGK 78
QY 72 YFELPENDQGWPRQPTSFKSEAYVDL-----TNEDANDTILEASPSG 115
Db 79 FFKGLASCGAW-----SCFDEFNRIDLEVLVSAQOILFTQRGINAGA-DTLLE-----G 128
QY 116 TPLEDSSSTIS-FITWNIDGLDGCNLPERARGVCSCLALYSPDVVFLQEVIPPYCAYLKR 174
Db 129 TEIKLDPTCSVFTTMNPGYAGRSDLPDNLKALFRIVAMMVPDYALISEIVLYSCGFIRAR 188
QY 175 AASYTIIT-----GNEGY-----FTAILLKGRVKFK---SQEIIPFPNTKMM 215
Db 189 PLSVKIVATYRLCSEQQLSSQHHYDYGMAVKSVLTAAGNLKLTKEPEDEDIL-----ML 242
QY 216 RNLCVNV 223
Db 243 RSINDVNL 250
RESULT 15
ID 09Y7M8 PRELIMINARY; PRT; 502 AA.
AC 09Y7M8;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Hypothetical 57.7 kDa protein.
GN SPBC9B6.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049769; CAB42372.1; -
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Hypothetical protein.
SQ SEQUENCE 502 AA; 57650 MW; 5ADFE2E7613B1A86 CRC64;

Query Match		5.1%;	Score 100.5;	DB 3;	Length 502;
Best Local Similarity		18.1%;	Pred. No. 1.9;		
Matches		88;	Conservative	61;	Mismatches 154; Indels 183; Gaps 18;
QY	62	DMOTOKALSAYPE-----LPE	NDGMPRPPTSFKSEAYVDLT	NEDANDTTILEAS	112
DB	22	DYRLSKLAISFFDQKKNMADSI	PVKKGSAGPP-SFVTPEYIEKQ	RQKKLEKMAKKA	80
QY	113	-----PSGTPLE-----	-----DSSTISFTWNIDGLD	CNL-----	139
DB	81	RKPIAPPSNAPPENTDFIKRE	MLSIPNYAPFETEKSA	LDITIMTYNV--LAQT	NIRRS
QY	140	-----PERARGVCSCLALYS	PDVYFLQEVIPYCA	YLKR---AASYTI--ITGN	184
DB	139	FPHSGEALKWKNSRMLANEL	TYYSPTLGCMQEVDAEFV	PNFYKKLLGGLGYELH	FIKGE
QY	185	EEGYFTAILLKKGRVKFKSQE	IIPF-----PNTKMRNL	LCV---NVS	LGNEFC
DB	199	GKTHGIMIFWKSLEK-KVQD	LTIYDDHDELLPGR	MNTKNIGCCVRLER	VDDPSRGLFLA
QY	234	TSHLESTREHSAERIRQ---	LKTVLGKQGEAPDST	VI	FAGDTN-----LRDQEV
DB	258	TTHLFWHPYGSYERLRQGA	ILVKEVNKAQSHPSW	PVFIAGDENTEP	PDNFPALTT
QY	281	IKCGGLPDNVFPAWEFL-G	KPKHCQYTWDTKAN	NN-----	-----
DB	318	SICORATDIERSMNYVFE	GESELEEKNA	STKTE	ENDSNEDDKECOS
QY	315	-----	-----LRIPAAYKH	RDRIF	---RAE
DB	378	KKRILHVQNDYVPHYRSFY	QHQEONPVLPSLY	SVGYKL	VHPENAKNTFDHPAFTNMAHAY
QY	334	EGHL-----	-----IPQSLDLV	GLEKLD	CGRFP
DB	438	QGHLDYIFVMNRDTSLOT	PENQVVEGIKL	KALLRVPLP	SEMKEAEPLE-GRYP
QY	364	CTLNVV	369		
DB	497	ANVQIV	502		

Search completed: November 14, 2002, 10:25:17
Job time : 30.1584 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 25.5928 Seconds
(without alignments)
2914.450 Million cell updates/sec

Title: US-09-697-863A-2
Perfect score: 1902
Sequence: 1 MELGSCLGEGREAEEEGEP.....DCGRFPSDHWGLCNLDIIL 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1902	100.0	362	4	O95551	O95551 homo sapien
2	1897	99.7	362	4	O9NXY9	O9NXY9 homo sapien
3	1893	99.5	362	4	O9NUK5	O9NUK5 homo sapien
4	1258	66.1	370	11	O9UJX7	O9UJX7 mus musculu
5	675	35.5	157	11	O9D7N0	O9D7N0 mus musculu
6	420	22.1	362	5	O9XWG3	O9XWG3 caenorhabdi
7	231	12.1	441	10	O9SA95	O9SA95 arabidopsis
8	127.5	6.7	449	5	O9VGS4	O9VGS4 drosophila
9	122.5	6.4	361	5	O8T2L4	O8T2L4 dictyosteli
10	113	5.9	502	3	O9Y7M8	O9Y7M8 schizosacch
11	111.5	5.9	252	16	O92X70	O92X70 rhizobium m
12	108.5	5.7	720	10	O9LSQ7	O9LSQ7 arabidopsis
13	108	5.7	253	16	O9S2L9	O9S2L9 streptomyce
14	107.5	5.7	863	15	O8UNX2	O8UNX2 human immun
15	106	5.6	267	16	O9CMW2	O9CMW2 pasteurella
16	106	5.6	1962	11	O9WUF3	O9WUF3 mus musculu

17	105.5	5.5	667	11	O8VCU0	O8VCU0 mus musculu
18	105	5.5	1374	10	O22148	O22148 arabidopsis
19	104.5	5.5	281	16	O8YPH2	O8YPH2 anabaena sp
20	103.5	5.4	256	16	O8YA41	O8YA41 listeria mo
21	103	5.4	288	5	O45711	O45711 caenorhabdi
22	103	5.4	1275	4	O9Y5K0	O9Y5K0 homo sapien
23	103	5.4	1275	4	O00366	O00366 homo sapien
24	102	5.4	265	16	O8UEP1	O8UEP1 agrobacteri
25	102	5.4	1275	4	O00375	O00375 homo sapien
26	101	5.3	405	5	O15921	O15921 trypanosoma
27	101	5.3	1275	4	O15604	O15604 homo sapien
28	101	5.3	1275	4	O12881	O12881 homo sapien
29	101	5.3	1275	4	O00368	O00368 homo sapien
30	101	5.3	1275	4	O00378	O00378 homo sapien
31	101	5.3	1275	4	O15606	O15606 homo sapien
32	101	5.3	1275	4	O8TE30	O8TE30 homo sapien
33	100.5	5.3	695	2	O85672	O85672 clostridium
34	100.5	5.3	816	9	O8SD65	O8SD65 pseudomonas
35	100	5.3	458	10	O9LVE3	O9LVE3 arabidopsis
36	100	5.3	1261	4	O15463	O15463 homo sapien
37	100	5.3	1275	4	O00370	O00370 homo sapien
38	100	5.3	1275	4	O9UN80	O9UN80 homo sapien
39	99.5	5.2	256	16	O8XLI6	O8XLI6 clostridium
40	98.5	5.2	283	5	O9BKP5	O9BKP5 dictyosteli
41	98	5.2	703	5	O8SW87	O8SW87 encephalito
42	98	5.2	1275	4	O00360	O00360 homo sapien
43	98	5.2	1275	4	O00362	O00362 homo sapien
44	98	5.2	3844	5	O94648	O94648 plasmodium
45	97.5	5.1	1001	11	O9JMC1	O9JMC1 rattus norv

ALIGNMENTS

RESULT 1
O95551 PRELIMINARY; PRT; 362 AA.
AC O95551;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE DJ30M3.3 (TTRAF and TNF receptor associated protein) (ETSL-associated protein 2) (Hypothetical 40.9 kDa protein).
GN DJ30M3.3 OR TTRAF OR EAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309820; PubMed=10764746;
RA Pyre S., Declercq W., Ibrahim A., Michiels C.,
RA Van Riettschoten A.G.I., Dewulf N., de Boer M., Vandenabeele P.,
RA Huylebroeck D., Remacle J.E.;
RT "TTRAF, a novel protein that associates with CD40, Tumor Necrosis Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs), and that inhibits Nuclear Factor-kappaB activation.";
RL J. Biol. Chem. 275:18586-18593(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Li R., Pei H., Papas T.S.;
RT "EAP2, a Novel Protein Interacting with ETSL.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031775; CAA21141.1; -.

DR EMBL: AJ269473; CAB92966.1; -
DR EMBL: AF201687; AAG35600.1; -
DR EMBL: BC017553; AAH17553.1; -
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
KW Receptor; Hypothetical protein.
SQ SEQUENCE 362 AA; 40929 MW; 37892E125DB64410 CRC64;

Query Match 100.0%; Score 1902; DB 4; Length 362;
Best Local Similarity 100.0%; Pred. No. 2e-150;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGREAAEEGEPEVKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNS 60
QY 61 YFEPVVEESALERPETISEPKTYVDLTNEETTDSTSKISPSEDTQOENGSMFSLITWN 120
Db 61 YFEPVVEESALERPETISEPKTYVDLTNEETTDSTSKISPSEDTQOENGSMFSLITWN 120
QY 121 IDGLDLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
Db 121 IDGLDLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
QY 181 AIMLKSRVKLKSQEIIFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
Db 181 AIMLKSRVKLKSQEIIFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
QY 241 LKMVLKKMQEAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGPKHKQYTWDTQ 300
Db 241 LKMVLKKMQEAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGPKHKQYTWDTQ 300
QY 301 MNSNLGITAACKLRFDRIFFRAAAEEGHIIPRSLDLGLGKLDGGRPSDHWGLLCNLDI 360
Db 301 MNSNLGITAACKLRFDRIFFRAAAEEGHIIPRSLDLGLGKLDGGRPSDHWGLLCNLDI 360
QY 361 IL 362
Db 361 IL 362

RESULT 2
Q9NYY9 PRELIMINARY; PRT; 362 AA.
AC Q9NYY9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE AD022 protein.
GN AD022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL GLAND;
RA Song H., Gao G., Peng Y., Ren S., Chen Z., Han Z.;
RT "A novel gene expressed in human adrenal gland.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF223469; AAF64144.1; -
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40956 MW; D6D607DE521C27E4 CRC64;

Query Match 99.7%; Score 1897; DB 4; Length 362;
Best Local Similarity 99.7%; Pred. No. 5.3e-150;
Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGREAAEEGEPEVKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNS 60

QY 61 YFEPVVEESALERPETISEPKTYVDLTNEETTDSTSKISPSEDTQOENGSMFSLITWN 120
Db 61 YFEPVVEESALERPETISEPKTYVDLTNEETTDSTSKISPSEDTQOENGSMFSLITWN 120
QY 121 IDGLDLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
Db 121 IDGLDLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
QY 181 AIMLKSRVKLKSQEIIFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
Db 181 AIMLKSRVKLKSQEIIFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
QY 241 LKMVLKKMQEAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGPKHKQYTWDTQ 300
Db 241 LKMVLKKMQEAPESATVIFAGDTNLRDREVTTCGGLPNNIVDVWEFLGPKHKQYTWDTQ 300
QY 301 MNSNLGITAACKLRFDRIFFRAAAEEGHIIPRSLDLGLGKLDGGRPSDHWGLLCNLDI 360
Db 301 MNSNLGITAACKLRFDRIFFRAAAEEGHIIPRSLDLGLGKLDGGRPSDHWGLLCNLDI 360
QY 361 IL 362
Db 361 IL 362

RESULT 3
Q9NUK5 PRELIMINARY; PRT; 362 AA.
ID Q9NUK5;
AC Q9NUK5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CDNA FLJ11306 fis, clone PLACE1010031.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK002168; BAA92119.1; -
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40869 MW; 8E1702C93BCB62E3 CRC64;

Query Match 99.5%; Score 1893; DB 4; Length 362;
Best Local Similarity 99.7%; Pred. No. 1.1e-149;
Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNS 60
Db 1 MELGSCLEGREAAEEGEPEVKRRLLCYEFASVASCDAVAQCFLAENDWEMERALNS 60
QY 61 YFEPVVEESALERPETISEPKTYVDLTNEETTDSTSKISPSEDTQOENGSMFSLITWN 120
Db 61 YFEPVVEESALERPETISEPKTYVDLTNEETTDSTSKISPSEDTQOENGSMFSLITWN 120
QY 121 IDGLDLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
Db 121 IDGLDLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNYEITGHEEGYFT 180
QY 181 AIMLKSRVKLKSQEIIFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240
Db 181 AIMLKSRVKLKSQEIIFPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQ 240

[illegible]

RESULT 4			
ID	Q9JUX7	PRELIMINARY;	PRT; 370 AA.
AC	Q9JUX7;		
DT	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE	Putative TRAF and TNF receptor associated protein.		
GN	TRAP OR TRAP.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20309820; PubMed=10764746;		
RA	Pype S., Declercq W., Ibrahim A., Michiels C.,		
RA	Van Rietschoten A.G.I., Dewulf N., de Boer M., Vandenabeele P.,		
RA	Huylebroeck D., Remacle J.E.;		
RT	"TRAP, a novel protein that associates with CD40, Tumor Necrosis		
RT	Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs),		
RT	and that inhibits Nuclear Factor-kappaB activation.";		
RL	J. Biol. Chem. 275:18586-18593(2000).		
DR	EMBL; AJ251328; CAB92971.1; -.		
DR	MGD; MGI:1860486; Ttrap.		
DR	InterPro; IPR005135; Exo_endo_phos.		
DR	Pfam; PF03372; Exo_endo_phos; 1.		
RW	Receptor.		
SO	SEQUENCE 370 AA; 41033 MW; A773A8889DF5BE83 CRC64;		

Query Match	66.1%;	Score 1258;	DB 11;	Length 370;
Best Local Similarity	68.4%;	Pred. No. 1e-96;		
Matches 240;	Conservative 43;	Mismatches 64;	Indels 4;	Gaps 3;

QY	12	EAAEEGEPEYKRRLLCVFEFASVASCDAVAQCFLAENDMEMERALNSYFEPPEVESAL	71
Dd	24	EAAQAE - EDVKRRLRLQLGCFALVGCCDPTMWPVSYLRENDWOTOKALSAYFELPENDGW	82
QY	72	ERRPETISEPKTYVDLTNEETTDSTTSKISPSEDTQOENGSMESLITWNIDGLDNMISE	131
Dd	83	PROPTSFKSEAYVDLTNEDANDTTILEASPS - GTPLEDSSITSFITWNIDGLDCNDPE	141
QY	132	RARGVCXYALYSPDVIFLOEVIPYYSYLKRRSSNYEITGHEEGYFTAIMLKRSRYKL	191
Dd	142	RARGVCSCALYSPDVVFLOEVIPYCAYLKKRAASYTIITGNEEGYFTAILLKGRKYF	201
QY	192	KSOEIIPPESTKMNRNLGVHVNVSNGNELCLMTSHLESTRGHAAERMNQLMVLKKMOEA	251
Dd	202	KSOEIIPENTKMNRNLLCVNVSLSGNEFCIMTSHLESTRSHAERIRQLKTVLGKMOEA	261
QY	252	PESATVTFAGDTNLRDREVTRCGGLPNNIIVDVWEFLGKPRKHCOYTWDTOMSNSNLGITAAC	311
Dd	262	PDSTVTI FAGDTNLRDQOEVIKCGGLP DN VFDAWEFLGKPKHC OYTWDTKANNMLRI PAAY	321
QY	312	KLRFDRIFFRAAAEEGHIIPRSLDLGLEKLD CGRPSPDHWG L LC N D I T L	362
Dd	322	KHRFDRIEFR - AEEGHLIPQSLDI VGLEKLD CGRPSPDHWG L LC T L NV L	370

RESULT	5
Q9D7NO	
ID	Q9D7NO
AC	Q9D7NO;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Traf and Tnf receptor associated protein.
GN	TTRAP.
OS	Fus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX	NCBI_TaxID=10090;

RC SEQUENCEFROM N.A.
RP STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Areakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009089; BAB26063.1; -
DR MGD; MGI:1860486; Ttrap.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 157 AA; 17876 MW; 5920850802FAAE84 CRC64;

Query Match	35.5%	Score 675;	DB 11;	Length 157;
Best Local Similarity	77.4%;	Pred. No. 1.2e-48;		
Matches 123; Conservative	18;	Mismatches 16;	Indels 2;	Gaps 1;

```
QY 204 MMRNLLCVHVNVSNGNELCLMTSHLESTRGAAERMÖLKMVLKKÖEAPESATVIFAGDT 263
|||||::: ||| |||||||||::: ||| ||| ||||||: |||||||
Db 1 MMRNLLCVNVSLSGNEFCMLTSHLESTRHSAERIRÖLKTVLGKÖEAPDSTVIFAGDT 60
|||||::: ||| |||||||||::: ||| ||| ||||||: |||||||
QY 264 NLRDREVTRCGGLEPNNIVDVWEFLGKPKHCÖYTWDTÖMNSNLGITAAKLRFDRIEFFRAA 323
|||||::: |||||||||: | ||||||||| |||||::: ||| ||| |||||||
Db 61 NLRDQEVIKCGGLPDNVFDAWEFLGKPKHCÖYTWDTKANNNLRIPAAKYKRFDRIFFR-- 118
|||||::: |||||||||: | ||||||||| |||||::: ||| ||| |||||||
QY 324 AEEGHIIPRSLDLGLLEKLDGGRFPSDHMGLLCNLDITL 362
|||||::: ||||||||| ||||||||| ||::: |||
Db 119 AEEGHLIPÖSLDLVGLEKLDGGRFPSDHMGLLCTLNVVL 157
```

RESULT 6		
Q9XWG3		
ID	Q9XWG3	PRELIMINARY; PRT; 362 AA.
AC	Q9XWG3;	
DT	01-NOV-1999 (TREMBLrel. 12, Created)	
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)	
DE	Y63D3A.4 protein.	
GN	Y63D3A.4.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	

```
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; Pubmed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
   investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL032652; CAA21707.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40893 MW; 7895C47504866E52 CRC64;

Query Match      22.1%; Score 420; DB 5; Length 362;
Best Local Similarity 29.3%; Pred. No. 7.1e-27;
Matches 108; Conservative 75; Mismatches 142; Indels 44; Gaps 10;

QY 7 LEGGREAEEF-----GEPEVKRRLLCY-----EFASVASCDAVAQCFLAEN 50
   :| | :| | | | | :| | | | | :| | | | | | | | | | | | | | |
Db 11 IEAKRQKMSQDSEVEIEILDEPEQKLNSSMSDEQKLHEFAITATDEAFAQSIQDV 70
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 51 DWEMERALNSYEFPVEESALERRPETISEPKTYVDLTNETTDTSTSKISPSEDTQEN 110
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71 DWDLKKALDVFY-----GSEAFAEARSAAVMGASSSSGAAVMTAEDLK--- 116
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 111 GSMFLITWNIDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNEYI 170
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 GFEVSVMSNIDGLDGRSLTRMKAVAHIVKNVPDILFLQEVVDRDLAPIDKLQSLYKI 176
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 171 ITGHEE-GYFTAIMLKSRVKLSQETIPEPSTKMRRNLGVHNVSGNELCLMTSHLES 229
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 YSNKGCQYTTAILVSK-MEDVEKHVDIHFQNSGMRYRTLQILEGSIIGLKVFLNTHLES 235
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 230 TRGHAERMNOLKMKVKKMQE---APESATVIFAGDTNLRDREVTTCGGLPNNIVDWEF 286
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 TREHRPQRCAGFCFCMDKVREIIAQNPGALVFFGGDLNLRDEEVS---VPDGVKDAWEA 292
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 287 LGGPKHCQYTDWTQMSNMLGITAACKLRFDRIFPRAAAEEGHIIPRSLDLGLEKL-DCG 345
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 293 AGSDNKTFTWDTFKNDNKQGFHGAKNRFDRLY----SGPLDKVKFTLEGRRIRISCL 347
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 346 RFPSDHWGL 354
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 348 CFPSDHWAI 356
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
Q9SA95 PRELIMINARY; PRT; 441 AA.
AC Q9SA95;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE F25C20.3 protein.
GN F25C20.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.M., Kremenetskaia I., Luros J., Ngan I., Liu A.,
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007296; AAD30241.1; -.
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR001876; znf_RangDP.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR SMART; SM00547; znf_RB2; 2.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS01358; ZF_RANBP2_1; UNKNOWN_1.
SQ SEQUENCE 441 AA; 48995 MW; 6582F9AA2D7BD0CA8 CRC64;

Query Match      12.1%; Score 231; DB 10; Length 441;
Best Local Similarity 30.2%; Pred. No. 5.3e-11;
Matches 84; Conservative 39; Mismatches 109; Indels 46; Gaps 11;

QY 113 MFSLITWNIDGLDNLNLSERARGVCSYLALYSPDVIFLQEVIPPYSYLKKR-----SS 166
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 175 ILSYNWFREDLELN---LRMRAIGHLIQLHSPHLICFQEVTPETIYDIFRKSNNWKAYSC 231
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 NYEITGHEEGYFTAIMLKSRVKLSQETIPEPSTKMRRNLGVHNVSGNE-LCLMTS 225
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 232 SVSVDVAVSRGYCYMLL--SKLGKVSFSKSGNSIMGRELIAIEVEVPGKKPLVFATS 288
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 HLES-----TRGHAERMNOLKMKVKKMQEAPESATVIFAGDTNLRDREVTTCGG-- 275
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 289 HLESPPGPPKWDQMFSSRERVEQAKFAEITLR---PNANVIFGDMWCD---KLDGKF 341
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 276 -LPNNIVDWEFLGKPKHCQYTDWTQMSNMLGITAACKLRFDRIFER----- 321
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 342 PLPDKVDVWEVL-KPGDLGFTYDTKANPMLSGNRALQKRLDRILCLDDYKLGIEWVG 400
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 322 AAEEGHIIPRSLDLG-LKLDGGRFSPSDHWGLCNL 358
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 401 KEAIPGLSYVKEKKVKGDIKKLELPVLPSPDHFGLLVTL 438
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
Q9VGS4 PRELIMINARY; PRT; 449 AA.
AC Q9VGS4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG4796 protein.
GN CG4796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
```

RA Fosler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003690; AAF54601.1; -.
DR FlyBase; FBgn0037872; CGA796.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 449 AA; 50275 MW; 1607DF25AA18A024 CRC64;

Query Match	6.78;	Score 127.5;	DB 5;	Length 449;
Best Local Similarity	20.08;	Pred. No. 0.023;		
Matches 74; Conservative	53;	Mismatches 108;	Indels 135;	Gaps 19;

[illegible]

RESULT 9		
Q8T2L4		
ID	Q8T2L4	PRELIMINARY; PRT; 361 AA.
AC	Q8T2L4;	
DT	01-JUN-2002 (TREMBLrel. 21, Created)	
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Hypothetical 41.3 kDa protein.	
OS	Dictyostelium discoideum (Slime mold).	
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium	
OX	NCBI_TaxID=44689;	

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Gulgo R., Kumpf K.,
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC115596; AAL92312.1; - .
 KW Hypothetical protein.
 DR EMBL; AC115596; AAL92312.1; - .
 SQ SEQUENCE 361 AA; 41327 MW; BBD46E1CE3B5135A CRC64;

Query Match	6.48;	Score 122.5;	DB 5;	Length 361;
Best Local Similarity	20.88;	Pred. No. 0.045;		
Matches 60;	Conservative 56;	Mismatches 93;	Indels 79;	Gaps 144;

[illegible]

```

RESULT 10
Q9Y7M8
ID Q9Y7M8 PRELIMINARY; PRT; 502 AA.
AC Q9Y7M8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 57.7 kDa protein.
GN SPBC9B6.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP
RC SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases
DR EMBL; AL049769; CAB42372.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Hypothetical protein.
SQ SEQUENCE 502 AA; 57650 MW; 54DFE2E7613B1A86 CRC64;

```

Query Match	5.98;	Score 113;	DB 3;	Length 502;
Best Local Similarity	18.28;	Pred. No. 0.44;		
Matches	90;	Conservative	65;	Mismatches 142;
				Indels 198;
				Gaps 20;

```
QY 51 DWEMERALNSYEP-----PVEESALERRPETISEPKYVDLTNEETDSTSK-- 99
      | : : | : | : : | : : | : : | : : | : : | : : |
Db 22 DYRLSKLAISFQDKNNKMADSI PVKKKGSAKGPSPFVTP- YIEKORQKKLEKMAKAA 80
      | : : | : | : : | : : | : : | : : | : : | : : |
QY 100 ---ISPSEDWQOENGSMF-----SLTWNIDGLDLNLT----- 129
      | : | : | : | : : : : | : : : : | : : : : |
Db 81 RKPIAPPSNAPENNTDFIKREMLSI PNPAPFETEKSA LDITMTYNNV--LAQTNI RSM 138
```


QY 130 -----SERARGVCSYLALSPDIVFLOEV-----IPPYSYLKRRSSNYEITGHE 175
Db 139 FPHSGEALKWKNNRSMLANELTYISPTLGCMQEVDAEFVNPFFYKKL-LGGLGYELHFIKG 197
QY 176 EGYFTAIML-KKSRVKLKSOEIIFF-----PSTKMMRNLCV----HVNVSQNELCLM 223
Db 198 EGKTHGIMLFWKSSLFKKVQDLTIYYDDHDELPGRMNTKNIGCCVRLERVDPPSRGLFLA 257
QY 224 TSHLESTRGHAERMNQLKMWLK--KMQAPESATVIFAGDTN-----LRDREV 270
Db 258 TTHLFWHPYGSYERLRQGAIIYKVEYNKMAQSHSPWPVEIAGDFNTEPEDTNFPALTTTPL 317
QY 271 TRCGGLPNNTVDVWEFLGPKHCQYTW-----DTQMSNLGITAAC----- 311
Db 318 SIC---QRAVDIIE-----RSMNVYGESELEEKNASTKTENDSNEDKKECQSSSTSS 368
QY 312 ----- 318
Db 369 VPESTASTPKKRILHVQNDYVPHYRSFYQOHEQNPLFLSLYSVGYKLVHPENAKNTEFDHP 428
QY 319 FFR--AAAEEGH-----IPRSLDLGLEKIDCGR 346
Db 429 AFTNWAHAYQGHLDYIFVMNRDTSLQTPDENQVEGIKLKALLRVPDPSEMKEAEPLF-GR 487
QY 347 FPSDHWGLCNLDII 361
Db 488 YPSDHALMANVQIV 502

RESULT 11

Q92X70 PRELIMINARY; PRT; 252 AA.
AC Q92X70;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein RB0092.
GN RB0092 OR SMB20092.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoeiter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603642; CAC48492.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 252 AA; 27376 MW; A8B6F28C8A8316AE CRC64;

Query Match 5.9%; Score 111.5; DB 16; Length 252;
Best Local Similarity 24.4%; Pred. No. 0.22;
Matches 61; Conservative 36; Mismatches 96; Indels 57; Gaps 12;

QY 136 VCSYLALSPDIVFLOEV-----IPPYSYLKRRSSNYEITGHEGYFTAIM 183
Db 34 IAAVIAECQPDVIALQEVGRARTGIDQAHMIAITHLMNEAEFHPALHLEDEKYYGDAVL 93
QY 184 LKKSRYKLKSOEIIFFPSTKMMRNLCVHVNVSQNELCLMTSHLESTRGHAERMNQLKM 243
Db 94 ---TALPMRLIKAPLPSSSEPRGALWEIDVAAVKQIVIVTHL-GLRG--AERLRQATA 147
QY 244 VLKK---MQEAPESATVIFAGDTNLRDREVTTRCGGLPNNTVDVWEFLG-KPKHCQYTWDT 299

Db 148 LLGPGWLGMAQGDHNVLAGDLNATGRS-TAYRLLARQLSDAQLLTGVKPR-----P 199
QY 300 QMNSNLGITAAACKLRFDRIFFRAAAEEGHIIPRSLDLGLEKIDC-----GRFPSDHW 352
Db 200 TFP SRLPL-----LRID-----HYLVGK---GIEVASCRVHGSTLARSASDHL 239
QY 353 GLLCNLDIIL 362
Db 240 PLLAELDVVM 249

RESULT 12

Q9LSQ7 PRELIMINARY; PRT; 720 AA.
AC Q9LSQ7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Genomic DNA, chromosome 5, BAC clone: F24B18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026634; BAA97521.1;
SQ SEQUENCE 720 AA; 81852 MW; 65B822BCF991672C CRC64;

Query Match 5.7%; Score 108.5; DB 10; Length 720;
Best Local Similarity 22.9%; Pred. No. 1.7;
Matches 78; Conservative 49; Mismatches 134; Indels 79; Gaps 16;

QY 8 EGGRAAEEGEPEV-----KKRLLCVFEFASVAS----- 37
Db 196 DGFRETREREGIPDLPEPESDYNLIRKNKKKKKKKNVRESSVASSEIDKRDEYANT 255
QY 38 CDAAVAQCFLAENDWEMERALNSYFEPVVEESALERRPETISEPKTYVDLTNEETDSTT 97
Db 256 CNGQVAD---ETD-----NSN-EACAQET--ENTPEIVTEVTRSD--EVEEVAR 298
QY 98 SKISPEDTQOENGSMESLITWNIDGL-DLNNLSERARGVCSYLALYS--PDVIFLOEVI 154
Db 299 SDEEVSDAEYESSSCFSESSGS--GLTDLRKVERINSICGAAGNSEVSELEVSRRV 356
QY 155 PPYYSYLKRRSSNYEITGHEGYFTAIMLKSRVKLKSOEIIFFPSTKMMRNLC--- 210
Db 357 --YHQPLGSQFKGFASRVIGSSGTTTRDLVLKRRFRIDDLAV---SLSMTEKLYMWEK 410
QY 211 -VHVNVSQNELC-----LMTSHLESTRGHAERMNQLKMWLKMKQAPESAT 256
Db 411 KLHAEVTAEEKLRVAYEKAYKILNNDQNGAESSELYEAEVL--VKLHLSKVNVSVRAVE 468
QY 257 VIFAGDTNLRDREVT-RCGGLPNNTVDVWEFLGPKHCQY 295
Db 469 SISMRIRKIRDEELSFOVIEITNGEFTWWRFLAKCHHKQF 508

RESULT 13

Q9S2L9 PRELIMINARY; PRT; 253 AA.
AC Q9S2L9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein SCO2125.
GN SCO2125 OR SC6E10.19C.

OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K., Harris D.;
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL109661; CAB51973.1; -
 DR InterPro; IPR005135; Exo_endo_phos.
 DR Pfam; PF03372; Exo_endo_phos; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 253 AA; 26691 MW; 1160CE2663C16622 CRC64;
 Query Match 5.7%; Score 108; DB 16; Length 253;
 Best Local Similarity 19.5%; Pred. No. 0.44;
 Matches 56; Conservative 52; Mismatches 119; Indels 60; Gaps 12;
 QY 95 STSKISPSSEDTQENGSMFSLITWNIDGLDNLSEARAGVCSYLALYSVDVIFLQEV 154
 DB 2 ATSLPALPNSRTEPDGSAVIRVLSYNI-----RSLRDDTDALARYIKACAPDLVLLQEA- 55
 QY 155 PPYYSYLK-----RSSNTEIITGHEGYFTAIMLK-KSRVKLSQEIIPFSTKMRNL 208
 DB 56 PRFRWRKKITRLAAGDVLVLSGGGTAGPALLCSLRATVERTEDVLLPLTPGHRRG 115
 QY 209 LCVHVNVSQNELCLMTSHLESTRGHAERMAOLKMKMOEAPESATVIFAGDTNLR-- 266
 DB 116 AAAYVRIGARIGVLSHSL---DADERHEQAQGLLLDHL-AALGVKHAVAGGDLNERPG 171
 QY 267 DREVTRCGGLPNIVDVWEFLGKPKHCQYTWDTQMSNLGITACKLRFDRIFFEAAAEE 326
 DB 172 GRTFRRLG---EGLRDCW--TAAPWGGEYTFP-----ATAPDRIIDAVFYTE---- 213
 QY 327 GHIPRSLDLGLEKIDCG-----RFPSDHWGLCNLDI 360
 DB 214 -----GIEVLGCGVPSGLAGVAEDDLRAATDHLPLVLTALRV 249
 RESULT 14
 O8UNX2 PRELIMINARY; PRT; 863 AA.
 ID O8UNX2;
 AC O8UNX2;

DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Envelope glycoprotein.
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MN;
 RX MEDLINE=21624264; PubMed=11752155;
 RA Zhang P.F., Bouma P., Park E.J., Margolick J.B., Robinson J.E.,
 RA Zolla-Pazner S., Flora M.N., Quinnan G.V. Jr.;
 RT "A variable region 3 (V3) mutation determines a global neutralization
 RT phenotype and CD4-independent infectivity of a human immunodeficiency
 RT virus type 1 envelope associated with a broadly cross-reactive,
 RT primary virus-neutralizing antibody response.";
 RL J. Virol. 76:644-655(2002).
 DR EMBL; AF443202; AAL66251.1; -
 DR InterPro; IPR00328; Env GP41.
 DR InterPro; IPR00777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 SQ SEQUENCE 863 AA; 98013 MW; EB6A2E57AA2F0A42 CRC64;
 Query Match 5.7%; Score 107.5; DB 15; Length 863;
 Best Local Similarity 21.2%; Pred. No. 2.7;
 Matches 66; Conservative 40; Mismatches 107; Indels 99; Gaps 13;
 QY 86 DLTNEETDST-----SKISPSSEDTQENGSM---FSLIT-----WN 120
 DB 132 DLRTNCTDSNTTCTNLGNTTNDNSKTTMEGEMKNCSEFNITTSIRDKMQKEYALYKLD 191
 QY 121 IDGLDNLNLSERAGV-CSYLALYSVDVIFLQEVIPY-----SYLK----- 162
 DB 192 IVAIDKDNISYRLISCNISYIQAQPKVSF--EPPIHYCAPAGFAILKCNKDNFTGKP 249
 QY 163 -KRSSNTEIITGHEGYFTAIMLKSRVKLSQEIIPFSTKMRN--LCVHVNVSQNE 219
 DB 250 CKNVSTVQCTHGIRPVVSTQLLNGS--LAEEVY-IRSENFNNAKTIIVHLNESVQI 305
 QY 220 LCLMTSHLESTRGH-----AARMNOLKMKMOEAPES 254
 DB 306 NCTRPYNRRRIHIGPGRAYTTKNIKGTIROAHCTISSAKWNPRLQIVSKLKEQFKN 365
 QY 255 ATVIFAGDTNLRDREVTRCGGLPNIVDVWEFLGKPKHC-----QYTWDTQMN 302
 DB 366 KTVFKQSS-----GGPEIVMHSFNGCGEYFYCNTSSLENSTWNGNNTWNNTTG 415
 QY 303 SNLGITACKLR 314
 DB 416 SNSNTTLQCKTK 427
 RESULT 15
 Q9CMW2 PRELIMINARY; PRT; 267 AA.
 ID Q9CMW2;
 AC Q9CMW2;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein PM0691.
 GN PM0691.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006105; AK02775.1; -.
DR HSSP; P09030; IAKO.
DR InterPro; IPR000097; Apendonclsel.
DR InterPro; IPR004442; ExoDNase_IIL.
DR InterPro; IPR004808; ExoIIL_xth.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR TIGRFAMS; TIGR00195; exoDNase_IIL; 1.
DR TIGRFAMS; TIGR00633; xth; 1.
DR PROSITE; PS00726; AP_NUCLEASE_F1_1; UNKNOWN_1.
DR PROSITE; PS00727; AP_NUCLEASE_F1_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 267 AA; 30756 MW; A1D50D1CC024D87 CRC64;

Query Match 5.6%; Score 106; DB 16; Length 267;
Best Local Similarity 20.7%; Pred. No. 0.7;
Matches 59; Conservative 46; Mismatches 112; Indels 68; Gaps 14;

QY 116 LITWINDGLDNLNLSERAGVCSYLALYSPIYFLQE--VIPPYSYLKKRSSNYEITG 173
DB 3 VISFNING----LRARPHQLEAVIEKYQPDVGLQEIKADEMFPELVEHLGYHYFHH 57
QY 174 HEEGYFTAIMLKSRVKLSQEIIPSPSTKMMRNLLCVHNVSGNELCLMTSHLE--STR 231
DB 58 GQKGHYGVALLTQAPK-AVRKGFPTDSEEAQKRIIMADLETDFGLLTVINGYFPQGESR 116
QY 232 GHAAERMNQLKV--LKKMQEAPESAT--VIFAGDTNLRDREVT-----R 272
DB 117 DHETKFPRAKEKYADLQQLLEODHNANNPIIIMGDMNISPTDLDIGDENRKRKWLRTGK 176
QY 273 CGGLPN-----NIVDWEFLGPKKHCQYTW---DTQMSNLSGITACKLRFD 317
DB 177 CSFLPEEREWMNRLYGYGLVDFRHLNPDVNDQFSWFDYRSKGFEDNRG-----LRIDH 230
QY 318 IFFRAA-----AEEGHIIPRSLDLGLEKLDGGRFSDH---WGL 354
DB 231 IIANHALAARCVETG---IALDIRAMEK-----PSDHAPVMAY 265

Search completed: November 14, 2002, 10:25:13
Job time : 28.5928 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 7.56388 Seconds
(without alignments)
2028.884 Million cell updates/sec

Title: US-09-697-863A-4
Perfect score: 1968
Sequence: 1 MASGSSSDAAEPAGPAGRAA.....DCGRFSDHGLCTLNVL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	129.5	6.6	388	1 NOCT_XENLA	P79942 xenopus lae
2	126	6.4	429	1 NOCT_MOUSE	O35710 mus musculu
3	123	6.2	431	1 NOCT_HUMAN	O9uk39 homo sapien
4	105.5	5.4	267	1 EX3_HAETN	P44318 haemophilus
5	99.5	5.1	556	1 PHL_LEPIN	P17627 leptospiro
6	98	5.0	494	1 PR31_YEAST	P49704 saccharomyc
7	98	5.0	1495	1 A2MG_MOUSE	O61838 mus musculu
8	93.5	4.8	1000	1 Y041_STYNY3	O55445 synechocyst
9	93.5	4.8	1246	1 SKIW_HUMAN	O15477 homo sapien
10	93	4.7	316	1 APE1_MOUSE	P28352 mus musculu
11	91	4.6	317	1 APE1_BOVIN	P23196 bos taurus
12	90.5	4.6	865	1 CYAA_PROMI	O59685 proteus mir
13	90	4.6	266	1 YAFD_ECOLI	P30865 escherichia
14	89.5	4.5	477	1 ISCI_YEAST	P40015 saccharomyc
15	89.5	4.5	663	1 VTER_HSV7J	P52462 human herpe
16	89	4.5	317	1 APE1_HUMAN	P27695 homo sapien
17	89	4.5	962	1 UBP4_MOUSE	P35123 mus musculu
18	88.5	4.5	1196	1 XPG_XENLA	P14629 xenopus lae
19	88	4.5	361	1 APEA_DICDI	P51173 dictyosteli
20	88	4.5	362	1 MFA3_HUMAN	P55082 homo sapien
21	88	4.5	549	1 G6PI_HAETN	P44312 haemophilus
22	87.5	4.4	354	1 ANGE_DROME	O24239 drosophila
23	87.5	4.4	422	1 NSMA_RAT	O9et64 rattus norv
24	87	4.4	342	1 QUEA_LISMO	O8y629 listeria mo
25	87	4.4	875	1 SYA_CHILMU	O9plh5 chlamydia m
26	86.5	4.4	1503	1 TRL2_HUMAN	O94759 homo sapien
27	85.5	4.3	4344	1 DYHC_EMENT	P45444 emericellia
28	85	4.3	316	1 APE1_RAT	P43138 rattus norv
29	85	4.3	342	1 QUEA_LISTIN	O92b13 listeria in
30	85	4.3	1135	1 RBL2_RAT	O55081 rattus norv
31	85	4.3	1391	1 RPOB_MYCPN	P78013 mycoplasma
32	84.5	4.3	476	1 PPBH_PSEAE	P35483 pseudomonas
33	84.5	4.3	550	1 G6PI_VIBCH	O9kuy4 vibrio chol

34	84.5	4.3	617	1 HEMA_MEASA	P35971 measles vir
35	84.5	4.3	2468	1 MAPB_HUMAN	P46821 homo sapien
36	84	4.3	353	1 ALCI_GORGO	P20758 gorilla gor
37	84	4.3	412	1 ACDS_RAT	P15651 rattus norv
38	84	4.3	573	1 C114_MOUSE	P19467 mus musculu
39	84	4.3	963	1 UBP4_HUMAN	O13107 homo sapien
40	84	4.3	1152	1 MAP4_HUMAN	P27816 homo sapien
41	83.5	4.2	536	1 ARP_ARATH	P45951 arabidopsis
42	83.5	4.2	978	1 PEX6_RAT	P54777 rattus norv
43	83.5	4.2	1017	1 LRGL_YEAST	P35688 saccharomyc
44	83.5	4.2	2700	1 ZAN_HUMAN	O9y493 homo sapien
45	83	4.2	353	1 DRN2_MOUSE	P56542 mus musculu

ALIGNMENTS

RESULT 1					
ID	NOCT_XENLA	STANDARD;	PRT;	388	AA.
AC	P79942;				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Nocturnin (Rhythmic message 1) (RM1).				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Retinal photoreceptors;				
RX	MEDLINE=97121484; PubMed=8962150;				
RA	Green C.B., Besharse J.C.;				
RT	"Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:14884-14888(1996).				
CC	-1- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED ONLY IN THE PHOTORECEPTORS OF THE RETINA. EXPRESSION IS CONTROLLED BY THE RETINAL CIRCADIAN CLOCK.				
CC	-1- SIMILARITY: BELONGS TO THE CCR4/NOCTURIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U74761; AAB39495.1; .				
DR	InterPro; IPR005135; Exo_endo_phos.				
DR	Pfam; PF03372; Exo_endo_phos; 1.				
KW	Biological rhythms.				
SO	SEQUENCE 388 AA; 43940 MW; 7C9F46F28D9C4E7E CRC64;				
	Query Match	6.6%;	Score 129.5;	DB 1;	Length 388;
	Best local Similarity	21.4%;	Pred. No. 0.00083;		
	Matches 72;	Conservative 50;	Mismatches 158;	Indels 57;	Gaps 14;
OY	74	ELPENOQGWPROPTSFKSEAYVLTNEDA--NDTITILEASPSGTPLDESSSTISFITWN	130		
DB	51	EASQHDQSEPLD-PKELLDECQVALQDRPARLHRDFPSLRSESSS---QDPRTRVMQWN	106		
OY	131	I-----DGLDG---CNL-----PERARGVCSCALYSPDYVFLQEV-----	163		
DB	107	ILAQALGEGKDNFIMCPMEALKWEERKYLILLEITLMQPDYVLCLOQEVDPHYFDTEQPILSR	166		
OY	164	IPPYCAYLKKRAASYTIITGNEGIFTAILLKKGRVKFKSQEIIIPPTNTKMRNLLCVNV	223		


```

DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE Sphingomyelinase C precursor (EC 3.1.4.12) (Sphingomyelin
DE sphingodiesterase) (SMASE).
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar Hardjo / isolate Sponselee;
RX MEDLINE=90307220; PubMed=2163985;
RA Segers R.P.A.M., van der Drift A., de Nijs A., Corcione P.,
RA van der Zeijst B.A.M., Gaastra W.;
RT "Molecular analysis of a sphingomyelinase C gene from Leptospira
RT interrogans serovar hardjo.";
RL Infect. Immun. 58:2177-2185(1990).
CC -! CATALYTIC ACTIVITY: Sphingomyelin + H(2)O = N-acylsphingosine +
CC choline phosphate.
CC -! SUBCELLULAR LOCATION: Secreted.
CC -! SIMILARITY: TO BACILLUS SPHINGOMYELINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52176; CAA36424.1; -.
DR PIR; S22634; S22634.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 556 SPHINGOMYELINASE C.
SQ SEQUENCE 556 AA; 63268 MW; 6D0EACDB9A5CD6FD CRC64;

```

Query Match	5.18;	Score 99.5;	DB 1;	Length 556;
Best Local Similarity	19.28;	Pred. No. 0.6;		
Matches	60;	Conservative	51;	Mismatches 106;
			Indels	95;
			Gaps	15;

QY	97	DLTNEDANDTILLEASPSGTPLEDS----	STISFITWNI-----	DGLDGCNLPERARG	145
		: : : : : : :	: :	:	
Db	55	DLTGGSVSSSPADAPENSILANSIPENMGIKILTHNVFLLEKTLPGWGNWGQNERAOR			114
QY	146	VCSCALYSPDVFLQEVIP-----	-----PYCAYLKRAASTYIITGNEEGY--		188
		: : : :	:	: : :	
Db	115	IVSSNYIQNDPIVFEDEAFDTPDARKILLDGVRESEYPY-----	-----QTDVIGRTKGMWA		165
QY	189	-----FT---ALLKKGVRKFSOEIIP-----	-----PPTKMMRNLLCVNVSISGN		228
		: : : : : :	: :	:	
Db	166	TLGLYRTDAFTNGCGVIVSVKWPPIEEKIQHVEKEKCGADVFESN---	KGFAVYRIDRNGR		221
QY	229	EFCIMTSHLESTREHSA----	ERIRQIKTVLG--KMQEAPDSTTVIFAGDTNL-----	RD	277
		: : : : :	: : :	: : :	:
Db	222	KFHIIGTHVQADSGCANLGVSVRVNQFNEIRDFIDSKKIPKHEMYLLIAGDLNVIKGSRE			281
QY	278	QEVIKCGGLPDNVEDAMWELGKPKHCQYTWDTKANNNLRIPAYAKHRFDRIFFRAEEGHL			337
		: : : : :	:	: :	
Db	282	YHQMCLILNVNNE--KYVGV--	FTWDTKTN--EIAAFYKKVE-----		319
QY	338	IPQSLDLVGLEK	349		
		: : :			
Db	320	PAYLDYIFVSK	330		

```

RESULT 6
PR31_YEAST
ID PR31_YEAST STANDARD; PRT; 494 AA.
AC P49704;
DT 01-OCT-1996 (Rel. 34, Created)

```

```

DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Pre-mRNA splicing factor PRP31.
OS PRP31 OR YGR091W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96184869; PubMed=8604353;
RA Weidenhammer E.M., Singh M., Ruiz-Noriega M., Woolford J.L. Jr.;
RT "The PRP31 gene encodes a novel protein required for pre-mRNA
RL Nucleic Acids Res. 24:1164-1170(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Wedler H., Scharfe M., Wedler E., Wambutt R.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PRE-MRNA SPLICING.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
CC -----
DR EMBL; U31970; AAA74984.1; -.
DR EMBL; Z72876; CAA97094.1; -.
DR SGD; S0003323; PRP31.
DR InterPro; IPR002687; Nop.
DR Pfam; PF01798; Nop; 1.
DR ProDom; PD004104; Nop; 1.
KW mRNA processing; mRNA splicing; Nuclear protein.
FT CONFLICT 361 394 TNPKRKELGENLGSTRKNSDYRMSDNKIGWNLA ->
FT DQPKKKRAGRGRKFRKYEKEKFLSHVRQLQNRMEFG (IN
FT REF. 2).
SQ SEQUENCE 494 AA; 55967 MW; 1C320B8630EEF5C8 CRC64;
```

Query Match	5.0%;	Score 98;	DB 1;	Length 494;
Best Local Similarity	20.6%;	Pred. No. 0.69;		
Matches 80;	Conservative 49;	Mismatches 116;	Indels 144;	Gaps 19;

```

OY      54 VPSVLRENDWQOTOKALSAYPELEBENDQGWRP-----PTSEKSEAYVDLTNEDAND 105
          | :| : : : | :|| : : ||| : : | :
Db     140 VISILENENYSKNESDELFFHL-ENKAKLTREQILVLTMSMKTSEFNKEPLDIKTR---- 194
          | :| : : : | :|| : : ||| : : | :
OY     106 TTILEASPSGTPLEDSSTISFTWNIDGLDGCNLPERARGVCSCLALYSPDVVFLOEVIIP 165
          | ||| : : || : : : : : : : : : | :| | |
Db     195 TQILEAN---SILEN-----LMKLQ-----EDIGQYASKISIAPNVCF--VGP 235
          | :| : : : | :|| : : ||| : : | :
OY     166 PYCAYLKRAAS-----YTITG-NEGYFTAILLKGRV 199
          | :| : : : | :|| : : ||| : : | :
Db     236 EIAQOLIAHAGGVLEFSRIPSCMIASIGKNKHLSHELHTLESVGVRQEGYL----- 285
          | :| : : : | :|| : : ||| : : | :
OY     200 KEKSQEITPEPNT--KMRRNLICVNVSLSGENEFLMTSHLESTREHSERIRQLKTVLG- 256
          | | | | : | : | : || | | | : : : : | :|
Db     286 -FASDMIQKFVPVSVHKQMLRMCAKAVSL-----AARVDAGQKNGDRNPVLAH 331
          | :| : : : | :|| : : ||| : : | :
OY     257 -----KMQEAP-DSTYIVAFADTNLRDQEVIKCG-----LPDNVEDAW 294
          | :| : : : | :|| : : ||| : : | :
Db     332 KWKAELS KARKKLEAPSISETKALPIPETPNRKRELGENLGSTRKNSDYRMSDNKIGW 391
          | :| : : : | :|| : : ||| : : | :
OY     295 EFLGPKRHCOYTWDT-----KANNNLRIPAAYKHREDRI 328
          | | : : : | :|| : : ||| : : | :
Db     392 NL---AKOEQTVLDSYGEEVGLGMSNTSLQDAVGATSGSRBSAGNOAKLTQVMKHRISEA 448
          | :| : : : | :|| : : ||| : : | :
OY     329 FFRAEE-----GHLLPQ---SLDLVGLEK 349
          | :| : : : | :|| : : ||| : : | :
Db     449 NQADEFLLSLGHNTQOPNLSPENYOQMHK 477

```

RESULT 7
A2MG_MOUSE STANDARD; PRT; 1495 AA.
AC Q61838; Q60628; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Alpha-2-macroglobulin precursor (Alpha-2-M).
GN A2M.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-44 AND 1240-1259.
RC TISSUE=Liver;
RX MEDLINE=93076803; PubMed=1280217;
RA van Leuven F., Torrekens S., Overbergh L., Lorent K., de Strooper B.,
RA van den Berghe H.;
RT "The primary sequence and the subunit structure of mouse alpha-2-
RT macroglobulin, deduced from protein sequencing of the isolated
RT subunits and from molecular cloning of the cDNA.";
RL Eur. J. Biochem. 210:319-327(1992).
RN [2]
RP SEQUENCE OF 1-161 FROM N.A.
RC STRAIN=129/J;
RX MEDLINE=95095249; PubMed=7528166;
RA Umans L., Serneeels L., Hilliker C., Stas L., Overbergh L.,
RA de Strooper B., van Leuven F., van den Berghe H.;
RT "Molecular cloning of the gene in embryonic stem cells."
RT and targeting of the gene in embryonic stem cells."
RL Genomics 22:519-529(1994).
CC -!- FUNCTION: IS ABLE TO INHIBIT ALL FOUR CLASSES OF PROTEINASES BY A
CC UNIQUE "TRAPPING" MECHANISM. THIS PROTEIN HAS A PEPTIDE STRETCH,
CC CALLED THE 'BAIT REGION' WHICH CONTAINS SPECIFIC CLEAVAGE SITES
CC FOR DIFFERENT PROTEINASES. WHEN A PROTEINASE CLEAVES THE BAIT
CC REGION, A CONFORMATIONAL CHANGE IS INDUCED IN THE PROTEIN WHICH
CC TRAPS THE PROTEINASE. THE ENTRAPPED ENZYME REMAINS ACTIVE AGAINST
CC LOW MOLECULAR WEIGHT SUBSTRATES (ACTIVITY AGAINST HIGH MOLECULAR
CC WEIGHT SUBSTRATES IS GREATLY REDUCED). FOLLOWING CLEAVAGE IN THE
CC BAIT REGION A THIOLESTER BOND IS HYDROLYZED AND MEDIATES THE
CC COVALENT BINDING OF THE PROTEIN TO THE PROTEINASE.
CC -!- TISSUE SPECIFICITY: PLASMA.
CC -!- DEVELOPMENTAL STAGE: CONTRARY TO THE RAT PROTEIN, WHICH IS AN
CC ACUTE PHASE PROTEIN, THIS PROTEIN IS ALWAYS PRESENT AT HIGH LEVELS
CC IN CIRCULATION.
CC -!- SIMILARITY: TO OTHER PROTEINS OF THE ALPHA-MACROGLOBULIN FAMILY,
CC INCLUDING COMPLEMENT COMPONENTS C3, C4, AND C5.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M93264; AAA39508.1; -
DR EMBL; U06977; AAA87890.1; -
DR HSSP; P01023; 1BV8.
DR MGD; MGI:87854; A2m.
DR InterPro; IPR002890; A2M_N.
DR InterPro; IPR001599; MacrogloblnA2.
DR Pfam; PF00207; A2M; 1.
DR Pfam; PF01835; A2M_N; 1.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
KW Serine protease inhibitor; Glycoprotein; Plasma; Bait region; Signal.
FT SIGNAL 1 24
FT CHAIN 25 1239 ALPHA-2-MACROGLOBULIN 165 KDA SUBUNIT.
FT CHAIN 1240 1495 ALPHA-2-MACROGLOBULIN 35 KDA SUBUNIT.
FT DOMAIN 686 744 BAIT REGION.

FT DISULFID 48 86 BY SIMILARITY.
FT DISULFID 249 298 BY SIMILARITY.
FT DISULFID 267 286 BY SIMILARITY.
FT DISULFID 277 430 BY SIMILARITY.
FT DISULFID 594 783 BY SIMILARITY.
FT DISULFID 642 689 BY SIMILARITY.
FT DISULFID 833 861 BY SIMILARITY.
FT DISULFID 859 895 BY SIMILARITY.
FT DISULFID 933 1339 BY SIMILARITY.
FT DISULFID 1092 1140 BY SIMILARITY.
FT THIOLEST 984 987 BY SIMILARITY.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1003 1003 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1385 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1443 1443 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 41 47 QESLKP -> SGIPEKA (IN REF. 2).
FT CONFLICT 91 91 I -> V (IN REF. 2).
FT CONFLICT 142 144 VKF -> GIT (IN REF. 2).
FT CONFLICT 160 161 FP -> VS (IN REF. 2).
SO SEQUENCE 1495 AA; 165827 MW; 0D215688E80B4CCB CRC64;

Query Match 5.0%; Score 98; DB 1; Length 1495;
Best Local Similarity 22.9%; Pred. No. 3.3;
Matches 66; Conservative 34; Mismatches 106; Indels 82; Gaps 15;

QY 26 AQAEDRVKRRRIQCLGFALVGG-----CDPTMVPVSLREND---WQTQ-----KALS 70
Db 1149 SQSQESHVYTKALAYAFALAGNKAKRSELLLESINKDAVKEEDSLHWQPGDVQKVALS 1208

QY 71 AYFELPENDQGWPRQPTSFKSEAYV--DLTNEDANDTTLLEASPSGTPLEDSSSTIS-F 126
Db 1209 FY-----QPRAPSAEVEMTAYVLLAYLTSESSRPTRLDSSS-----DLSTASKI 1252

QY 127 ITW-----NIDGLDCNLPERARGVCSCLALYSPDVFLQEVIPPCAYLKKRAASYTII 181
Db 1253 VKWISKQNSDQ-----GL-----LLTQDTVALQALSKYGSATFTRSQKEVLV 1296

QY 182 TGNBEGYFTAILLKGRVKFKSQEI--IP-----FPNTKMMRNLLCVNVS 224
Db 1297 TSSSGTFSTFHVNSGNRLLQLQEVRLPDLPGNYVTKSGSGCVYIQTSLKYNIL--PVA 1354

QY 225 LGGNEFCLMTSHL---ESTREHSAERIRQLKTVLGKMQEAPDSTTVI 268
Db 1355 DGAAPFALQVNTLPDLPNFDKAEDHRTFQIRINVSYTG---ERPSSNMVI 1399

RESULT 8
Y041_SYNY3 STANDARD; PRT; 1000 AA.
ID Y041_SYNY3
AC Q55445;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical methyl-accepting chemotaxis protein sl10041.
GN SL10041.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugitara M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";


```
RL DNA Res. 2:153-166(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; D64006; BAA10787.2; -
DR HSSP; P02942; 1QV7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR001294; Phytochrome.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF01590; GAF; 2.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00065; GAF; 2.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50046; PHYTOCHROME_2; 2.
KW Hypothetical protein; Transducer; Complete proteome.
SQ SEQUENCE 1000 AA; 108334 MW; 58C5A765E1275FDF CRC64;
```

Query Match 4.8%; Score 93.5; DB 1; Length 1000;
Best Local Similarity 19.8%; Pred. NO. 4.7;
Matches 78; Conservative 56; Mismatches 121; Indels 139; Gaps 18;

```
QY 1 MASGSSDAEPAGPAGRAASAPEAAQAEDRVKRRRLQCLGFALVGGCDPTWPSVLRE 60
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 17 VANGETLDGALFTGLDTAAQAODESSETS-----ASFATIDGEXKSEVGDL-- 63
QY 61 NDW-----QTQKALSAFFELPENDQGWPRQPTSFKSEAYV 96
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 64 -DMFSDGETDINGDEADRGIQADTQAKTLVSLNLTREEPEID-----GAFFMAEAFI 115
QY 97 DLTNEDANDTILEASPSGTPLEDSSTISFTWIDGL---DGCNLP-----RARGVCSC 149
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 116 -----AENTAVEDVSPNPNPAIDTALAALQSAVELTTPPPINLPKVELPPMQLAPL 169
QY 150 LALYSPDVVFLQEVIPYCAYLKKRAASYTITGNEGYFTAILLK-----GRVKFKS 203
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 170 MAIADPD-----NLSPMSTSIQAPTQSGSLRN-KAVLIALLIGLPAGVIGLNLSS 222
QY 204 QETIPEPNTKMRNLVCNVSLGNEFCLMTSHLESTREHSAERIRQLKTVLGMQEAPD 263
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 223 VDRLPVPQTE-----QQVKD 237
QY 264 STTVIFAGDTNLRDQEVITKCGGLPDNV---FDAMEFLGKPKHCQYTWDTKANNNLRIPAA 320
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 238 STT-----KQIRDQILI--GLLVTAVGAAFYAWVWGE-----NTKAQTALALKAK 281
QY 321 YKHR-FDR-IFRAEEGHLIPQSLDLVG--LEKL 350
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 282 HSHRNLDQPLAVAGDELAIDQITDALSAQVEKL 315
```

```
RESULT 9
SKIW_HUMAN
ID SKIW_HUMAN STANDARD; PRT; 1246 AA.
AC Q15477; Q12902; O15005; Q15476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Helicase SKI2W (Helicase-like protein) (HLP).
GN SKIV2L OR SKIV2 OR SKI2W OR DDX13 OR W.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95334363; PubMed=7610041;
RA Dangl A.W., Shen L., Mendoza A.R., Wu L.-C., Yu C.Y.;
RA Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT "Human helicase gene SKI2W in the HLA class III region exhibits
RT striking structural similarities to the yeast antiviral gene SKI2 and
RT to the human gene KIAA0052: emergence of a new gene family.";
RL Nucleic Acids Res. 23:2120-2126(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Albertella M.R., Jones H., Thomson W., Olavesen M.G.,
RA Campbell R.D.;
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-958 FROM N.A.
RC TISSUE=Hippocampus;
RA Lee S.-G., Song K.;
RT "Cloning of the human helicase-like protein gene homologous to Yeast
RT SKI2.";
RL Korean J. Biochem. 26:215-219(1994).
RN [5]
RP SEQUENCE OF 230-960 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95278935; PubMed=7759100;
RA Lee S.-G., Lee I., Park S.H., Kang C., Song K.;
RT "Identification and characterization of a human cDNA homologous to
RT yeast SKI2.";
RL Genomics 25:660-666(1995).
RN [6]
RP REVISIONS.
RA Song K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HELICASE; HAS ATPASE ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; Z48796; CAA88733.1; ALT_INIT.
DR EMBL; AF019413; AAB67978.1; -.
DR EMBL; X98378; CAA67024.1; -.
DR EMBL; U09877; AAB52523.1; ALT_INIT.
DR Genew; HGNC:10898; SKIV2L.
DR MIM; 600478; -.
DR InterPro; IPR001410; DEAD.
```



```

FT METAL 94 94 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 208 208 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 210 210 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 306 306 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 307 307 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT SITE 210 210 IMPORTANT FOR SUBSTRATE RECOGNITION (BY
SIMILARITY).
FT ACT_SITE 307 307 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 35359 MW; CF899FA8182AA558 CRC64;

Query Match 4.7%; Score 93; DB 1; Length 316;
Best Local Similarity 20.1%; Pred. No. 1;
Matches 48; Conservative 38; Mismatches 75; Indels 78; Gaps 10;

QY 83 PROPTSPKSEAYVDLTNEDA-----NDTILEASPSGTPLEDSSSTISFITWNIDGL 134
| : | : || : | : | : |||| : | : | : || : ||
DB 15 PKSEPETKKSKGAAKKTENKAAGEGPVLYEDPPDQKTSFSG----KSATLKICSWNV DGL 70

QY 135 DGCNLPERA---RGVSCIALYSPDVVLQEV-----IPPYCAYLKKRAASYTIITGN 184
| | : | : | : || : ||| : | | | : |
DB 71 -----RAWIKKKGL-DWVKEAPDILCLQETKCSNKLPABLQLPGLTHQYWSAPSD 122

QY 185 EEGYFTAILKKGRVKFSQEIIPEPNTKMNRLLCVNVSIGNE-----F 230
||| || : | : | : || : | : | : || : |
DB 123 KEGYSGVGLLSR-----QCPLKVSYGIGEEHHDQGRVIVAEFESEF 163

QY 231 CLMTSHLEST-----REHSERIRQLKTVLGKMQEAPDSTTVIFAGDTNLRDQEV 280
| : | : | : | : | : | : | : | : | : | : | : | : |
DB 164 VLVTA YVPNAGRGLVRLEYRQRWDEAFRRK-----LKDLASRKRPVLVLCGDLNVAHEEI 216

RESULT 11
APEL_BOVIN STANDARD: PRT; 317 AA.
ID APEL_BOVIN STANDARD: PRT; 317 AA.
AC P23196;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) (AP
endonuclease 1) (APEX nuclease) (APEN).
GN APEX OR APE OR BAPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
RC TISSUE=Thymus;
RX MEDLINE=91212189; PubMed=1708495;
RA Robson C.N., Milne A.M., Pappin D.J.C., Hickson I.D.;
RT "Isolation of cDNA clones encoding an enzyme from bovine cells that
repairs oxidative DNA damage in vitro: homology with bacterial repair
enzymes.";
RL Nucleic Acids Res. 19:1087-1092(1991).
RN [2]
RP SEQUENCE OF 1-22.
RC TISSUE=Thymus;
RX MEDLINE=87289027; PubMed=2441359;
RA Henner W.D., Kiker N.P., Jorgensen T.J., Munck J.-N.;
RT "Purification and amino-terminal amino acid sequence of an
apurinic/aprymidinic endonuclease from calf thymus.";
RL Nucleic Acids Res. 15:5529-5544(1987).
CC -1- FUNCTION: REPAIRS OXIDATIVE DNA DAMAGES IN VITRO. MAY HAVE A ROLE
IN PROTECTION AGAINST CELL LETHALITY AND SUPPRESSION OF MUTATIONS.
CC BREAKS THE BLOCKING GROUPS FROM THE 3' TERMINI OF THE DNA STRAND
CC CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
```

```

CC -1- TISSUE SPECIFICITY: THYMUS.
CC -1- INDUCTION: BY SEVERAL DNA DAMAGING AGENTS.
CC -1- SIMILARITY: BELONGS TO THE AP/EXO A FAMILY OF DNA REPAIR ENZYMES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, X56685; CAA40014.1; -.
DR PIR, S26830; S26830.
DR HSSP, P27695; 1HD7.
DR InterPro; IPR000097; Apendonclse1.
DR InterPro; IPR004442; ExoDNase_III.
DR InterPro; IPR004808; ExoIII_xth.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR TIGRFAMS; TIGR00195; exoDNase_III; 1.
DR TIGRFAMS; TIGR00633; xth; 1.
DR PROSITE; PS00726; AP_NUCLEASE_F1_1; 1.
DR PROSITE; PS00727; AP_NUCLEASE_F1_2; 1.
DR PROSITE; PS00728; AP_NUCLEASE_F1_3; 1.
DR DNA repair; Lyase; Nuclear protein.
FT INIT_MET 0 0
FT METAL 67 67 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 95 95 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 209 209 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 211 211 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 307 307 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT METAL 308 308 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT ACT_SITE 308 308 GENERAL BASE (BY SIMILARITY).
FT CONFLICT 20 21 PE -> LP (IN REF. 2).
SQ SEQUENCE 317 AA; 35438 MW; 58A7B2029B0891B4 CRC64;

Query Match 4.6%; Score 91; DB 1; Length 317;
Best Local Similarity 21.3%; Pred. No. 1.5;
Matches 51; Conservative 36; Mismatches 74; Indels 78; Gaps 11;

QY 83 PROPTSFKSEAYVDLTNEDA-----NDTTLLEASPSGTPLEDSTISFTIWNIDGL 134
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 16 PKTEPEAKSKSAGAKKNEKEAVGEGAVLYEDPDQKTSPSG-----KSATLKICSMVNDGL 71

QY 135 DGCNLPERA-----RGVCSCLALYSPDVVFLQEV-----IPPYCAYLKKRAASYITITGN 184
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 72 -----RAWIKKKGL-DWVKEEAPDILCLQETKCSENKLPVELQELSGLSHQYWSAPSD 123

QY 185 EEGYFTAILKKGRVKFKSQEIIFFPNTKMMRNLLCVNVS LGNE-----F 230
   : || | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 124 KEGYSGVGLSR-----QCP LKVSYGIGEBEHDOEGRVIAEYDAF 164

QY 231 CLMTSHLEST-----REHSAERIRQLKTVLGKMQEAPDSTTVIFAGDTNLRDOEV 280
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 165 VLVTAYVPNAGRGIVRLLEYRQRWDFAFR--KFLKGLASRRP----LVLCGLDNVAHEEI 217

RESULT 12
CYAA_PROMI STANDARD; PRT; 865 AA.
AC Q59685;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenyllyl
DE cyclase).
DE CYA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]

```

```

RP. SEQUENCE FROM N.A.
RX MEDLINE=97028791; Pubmed=8874804;
RA Trotot P., Sismeiro O., Vivares C., Glaser P., Bresson-Roy A.,
RL Danchin A.;
RT "Comparative analysis of the cya locus in enterobacteria and related
RT Gram-negative facultative anaerobes.";
RL Biochimie 78:277-287(1996).
CC CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -|- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-1 FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outpost on
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to licens@isb-sib.ch).
CC CC -----
CC CC EMBL; U22969; AAC44330.1; -.
DR DR InterPro; IPR000274; Adenylt_cyclase_1.
DR DR Pfam; PF01295; Adenylate_cyclia; 1.
DR DR PROSITE; PS01092; ADENYLATE_CYCLASE_1_1; 1.
DR DR PROSITE; PS01093; ADENYLATE_CYCLASE_1_2; 1.
KW KW Lyase; CAMP synthesis.
FT FT DOMAIN 1 540 CATALYTIC (POTENTIAL).
FT FT DOMAIN 546 865 REGULATORY (POTENTIAL).
SQ SQ SEQUENCE 865 AA; 99449 MW; 99D4B1D51FCB7D7 CRC64;

Query Match 4.6%; Score 90.5; DB 1; Length 865;
Best Local Similarity 21.8%; Pred. No. 7;
Matches 49; Conservative 25; Mismatches 52; Indels 99; Gaps 12;

QY 152 LYSPPDVVFLQEVIIPRYCA-----YLKRRASYYITIGNEE-----GYTAIL 193
| | : | : : | | | | | | : | | : | |
DB 448 LSEPHLETFI--YVPGRANRSGWLYLNRAPDFAHIVGHQPLEYNRYLNKLVAWSYFNGLL 505
| | : | : : | | : | | : | | : | |
QY 194 LKKGRV-----KFKSQELI-----FPNTKMM-----RNL-LCVNV-- 223
| : | : | : | : | : | : | : | : | : |
DB 506 TKDSQVYIHQGDDSCDEIKLHELVRDISHFPIRLPAPTPKALYSPCEIRHLAIIIVNLEV 565
| : | : | : | : | : | : | : | : | : |
QY 224 -----SLGNEFCLM-----TSHLESTREHSAER 247
| | : | | | : | | : | | : | : |
DB 566 DPTREFSDOVVHFDFRKLDVFESFGEEEOCLLGISTDLLRNSMWNEVRTLHFNGTQ---SM 621
| | : | | | : | | : | | : | : |
QY 248 IROLKVTLGKMGE--APPDSTVIFAGDTNLR-----DQEVIKC 283
: ||||| : | : : | : | : | : |
DB 622 LESLKTIIGKMHDAAAPPASVEVFCYSQHRLGLIRTRVQQLVSEC 666
| : | : | : | : | : | : | : | : | : |

RESULT 13
YA_FID_ECOLI STANDARD; PRT; 266 AA.
AC P30865; P75671;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yaF_D.
GN YA_FID OR B0209 OR Z0232 OR ECS0205.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nishimura K., Komine Y., Miyamoto K., Kitabatake M., Mathunaga F.,
RL Hisano T., Miki T., Inokuchi H.;
RN Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
```

```

RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hataeda E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RL 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RN Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RN Nature 409:529-533(2001).
RL [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RL O157:H7 and genomic comparison with a laboratory strain K-12.";
RN DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 47-259 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92137666; PubMed=1663890;
RA Kajle S.-I., Ideta R., Yamato I., Anraku Y.;
RT "Molecular cloning and DNA sequence of dhfr, a gene affecting
RL anaerobic expression of the Escherichia coli hexaheme nitrite
RT reductase.";
RN FEMS Microbiol. Lett. 67:205-211(1991).
RN -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D12650; BAA02172.1; ALT_INIT.
DR EMBL; AE000130; AAC73314.1; -.
DR EMBL; D83536; BAA77880.1; ALT_INIT.
DR EMBL; U70214; AAB08631.1; ALT_INIT.
DR EMBL; AE005196; AAG54505.1; -.
DR EMBL; AP002550; BAB33628.1; -.
DR EMBL; X60739; -. NOT_ANNOTATED_CDS.
DR PIR; JS0717; JS0717.
DR EcoGene; EG11650; yafD.

```



```

DR InterPro;IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 112 112 T -> S (IN REF. 7).
FT CONFLICT 223 223 R -> G (IN REF. 7).
SQ SEQUENCE 266 AA; 29991 MW; 5F5AD2B0C4CE5794 CRC64;

Query Match 4.6%; Score 90; DB 1; Length 266;
Best Local Similarity 20.6%; Pred. No. 1.4;
Matches 51; Conservative 20; Mismatches 55; Indels 122; Gaps

QY 81 GWPQRQ--PTSEKSEAYVDLTNEDANDTILEASPSGTPLEDSSTISFITWNI----- 131
Db 13 GQPAERILPPGSFAS-----IGQALPGEPLSTERIRILWNIIYKQRA 57
QY 132 -----DGLDGNLPERARGVCSCLALY 153
Db 58 EWLVLKNGKDAHLVLLQEAQTPELVQFATANYLAADQVPAFVLPQHPSCVMTLSAAH 117
QY 154 SPDVFLQEVIPPCAYLK-----RASVTITGNEGYFTAILLKGKRVKFKSQEILP 208
Db 118 -----PVYCCPLREREPILRLAKSALVT-----VYP 143
QY 209 FPNTRK--MMRNLLCVNVLGNEFCIMTSHLESTREHSAERIRQLKTVLGKMQEAPDST 266
Db 144 LPTRLMNVNIHAVNFSLGVDVYS-----KQLPLI--GDQIAHHS GP 184
QY 267 VIFAGDTN 274
Db 185 VIMAGDFN 192

RESULT 14
ISCL_YEAST
ID ISCL_YEAST STANDARD; PRT; 477 AA.
AC P40015;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Inositol phosphosphingolipids phospholipase C (EC 3.1.4.-) (IPS
DE phospholipase C) (IPS-PLC) (Neutral sphingomyelinase) (N-Smase)
DE (nSmase).
GN ISCL OR YER019W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RN
RP CHARACTERIZATION.
RP MEDLINE=20564359; PubMed=11006294;
RX Sawai H., Okamoto Y., Luberto C., Mao C., Bielawska A., Domae N.,
RA Hannun Y.A.;
RA "Identification of ISCL (YER019W) as inositol phosphosphingolipid
RT phospholipase C in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 275:39793-39798(2000).
CC -|- FUNCTION: Responsible for the hydrolysis of the
CC phosphosphingolipids (IPS), inositol phosphorylceramide (IPC),
CC mannosylinositol phosphorylceramide (MIPC), and mannosyldiinositol
CC phosphorylceramide (M(IP)2C). Also active on sphingomyelin, but
CC this activity is probably not physiologically relevant.
CC -|- COFACTOR: Magnesium.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.

```

[illegible]

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
DR EMBL; U043400; AAC54723.1; -.
DR InterPro; IPR003498; DNA_pack_C.
DR InterPro; IPR003499; DNA_pack_N.
DR Pfam; PF02499; DNA_pack_C; 1.
DR Pfam; PF02500; DNA_pack_N; 1.
KW DNA packaging.
SQ
SEQUENCE 663 AA: 76411 MW: 800206FA1D166DEC CRC64:
```

Query Match	4.5%;	Score 89.5;	DB 1;	Length 663;
Best Local Similarity	20.2%;	Pred. No. 5.8;		
Matches	74;	Conservative 38;	Mismatches 125;	Indels 129;
			Gaps	17;

QY	49	CDP-----TWPSVLREN-DWQOTKALSAVFELPENDQGWBRQPTSEKSEAYV----	96
		::::	
Db	79	CDPLNCALSKLKEFTAIIEKNTVEYQKHLELQTSFYRN-----PMELQIEKEFIQDFQ	130
QY	97	-----DLTNEDANDTTLLEASPSGTPLEDSDSTISFITWNIDGLDGCNLPERARGVCSCLA	151
		: : : : :	
Db	131	RWICGDFENTNKKERIKLE-----PFQKSILIHIEF-----ISVTKLPTLANHVLDYLK	180
QY	152	LSPSDVVFLOE-----VIPPYCAVLKRAASYTIITGNEEG	187
		: : : :	
Db	181	-YKFDIEEFINESSVNILKOKASVFLVPRRHGKTFWMIPVICFLKN-----LEGISIG	232
QY	188	Y-----FTAILLKGRVKFKSOEIIPEPNTKMMRNL-----CVNV-SL	225
		: : : : : :	
Db	233	YVAHQKHVSHFWMKD--VEFKCRREFPOKNITCODNVITIEHTIKSTALFASQYNTHSI	290
QY	226	GGNEFCIM---TSHLESTREHSAERIQLKTVLGKMQEAPDSTTVIFAGDTNLRDQEYIK	282
		: : : : : : :	
Db	291	RGOSFNLLIVDESHFIKKDAFS-----TILGFLPOS-STKIIFISSTNSGNHSTSF	340
QY	283	CGGLPDNVFD-----AMEFLGKPKHCQYTWDTKANNNLRI	317
		: : : :	
Db	341	LTKLSNSPEMELTVVSYVCEDHVIHLNDRGNATTACACYRLHKPKFISINADVKKTDLFL	400
QY	318	PAAYKH 323	
		:	
Db	401	EGAFKH 406	

Search completed: November 14, 2002, 10:25:48
Job time : 10.5639 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 ; Search time 13.8671 Seconds
(without alignments)
2565.041 Million cell updates/sec

Title: US-09-697-863A-4
Perfect score: 1968
Sequence: 1 MASGSSSDAAEPAGPAGRAA.....DCGRFPSDHWGLCTLNVVL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	432.5	22.0	362	2	T27272	hypothetical prote
2	275	14.0	441	2	B86252	hypothetical prote
3	118.5	6.0	253	2	T35506	hypothetical prote
4	111.5	5.7	252	2	D95853	conserved hypothet
5	105.5	5.4	267	2	H64044	exodeoxyribonuclea
6	102	5.2	265	2	F97566	probable nuclease
7	102	5.2	265	2	AD2787	exonuclease III [i
8	101	5.1	1114	2	T30299	dynein heavy chain
9	100.5	5.1	502	2	T40792	hypothetical prote
10	99.5	5.1	556	2	S22634	sphingomyelin phos
11	99.5	5.1	678	2	H96552	hypothetical prote
12	98	5.0	1495	2	S27001	alpha-2-macroglobu
13	97	4.9	2218	2	B84683	hypothetical prote
14	96.5	4.9	840	2	S74707	nitrogen fixation
15	94.5	4.8	549	2	T02790	hypothetical prote
16	94	4.8	1240	2	T48800	SMT4 related prote
17	93.5	4.8	282	2	T51013	related to calmodu
18	93.5	4.8	1246	2	S56752	helicase SKI2W - h
19	93	4.7	270	2	H83327	exodeoxyribonuclea
20	93	4.7	317	2	A39500	DNA-(apurinic or a
21	92.5	4.7	288	2	T24066	exonuclease III ho
22	91	4.6	318	2	S26830	DNA-(apurinic or a
23	90	4.6	266	2	C64745	yafD protein - Esc
24	90	4.6	266	2	E90654	yafD protein [simi
25	90	4.6	266	2	E85505	yafD protein [simi
26	90	4.6	608	2	G87638	methyl-accepting c
27	90	4.6	691	2	C84300	cell division cycl
28	89.5	4.5	477	2	S50477	hypothetical prote
29	89.5	4.5	663	2	T41963	hypothetical prote

30	89.5	4.5	750	2	AB0708	catalase (EC 1.11.
31	89.5	4.5	923	2	A86334	T20H2.17 protein -
32	89	4.5	318	2	S23550	DNA-(apurinic or a
33	89	4.5	571	2	H97070	membrane associete
34	88.5	4.5	838	2	G84599	probable retroelem
35	88.5	4.5	1196	2	S35994	DNA repair protein
36	88	4.5	290	2	AC3417	DNA polymerase, ba
37	88	4.5	361	2	S68268	microfibril-associ
38	88	4.5	362	2	A56745	glucose-6-phosphat
39	88	4.5	563	2	F64130	geneC protein [imp
40	87.5	4.4	536	2	AH3227	conserved hypothet
41	87	4.4	259	2	AH0533	S-adenosylmethioni
42	87	4.4	342	2	AC1266	probable aldehyde
43	87	4.4	497	2	B82981	alanyl-tRNA synthe
44	87	4.4	875	2	H81739	Mutator-like trans
45	86.5	4.4	435	2	G84518	

ALIGNMENTS

RESULT 1
T27272
hypothetical protein Y63D3A.4 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T27272
R/White, S.
submitted to the EMBL Data Library, October 1998
A/Reference number: Z20335
A/Accession: T27272
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-362 <WIL>
A/Cross-references: EMBL:AL032652; PIDN:CAA21707.1; GSPDB:GN00019; CESP:Y63D3A.4
A/Experimental source: clone Y63D3A
C/Genetics:
A/Gene: CESP:Y63D3A.4
A/Map position: 1
A/Introns: 18/1; 43/1; 116/3; 148/2; 225/3; 320/2

Query Match 22.0%; Score 432.5; DB 2; Length 362;
Best Local Similarity 33.5%; Pred. No. 2e-28;
Matches 113; Conservative 55; Mismatches 126; Indels 43; Gaps 11;

QY	43	FALVGGDPFWVPSVLR	ENDMOTOKALSAYFELLPENDQGWPRDPTSFKSEAYVDLTNE	D	52	FALITATDEAFAGSILQDV	WDLKKALDYF-----YGSEAF	A-----E	89
QY	103	ANDTILEAS-----	PSGTPL--ED--S	TISFITWNIDGLDGCNLLPERARGVCSCLALY	153				
Db	90	ARSAAVMGASSSSMASSGA	AVMTAEDLKGEFVSVM	SWNIDGLDGRSLITRMKAVAHYKNV	149				
QY	154	SPDVFLQEVIPYCAYLK	KRAASTYIIITGNEE-GYFTAILKKGRVKFSQETIPPE	NT	212				
Db	150	NPDIQLFQEVYVDRDLA	PTDKLSLYKIIYSNKG	CQYTTALVSK-MFDVEKHDIHFNQNS	208				
QY	213	KMRNLLCVNVSLGNEF	CLMTSHLESTREHSAERIRQLKTVLGKMQE--APDSTVI	F	269				
Db	209	GMVRTLQILEGSIGLK	VFLNLNTHLESTREH	RPQRCACFGFCMDKVR	EIIAQNP	GALVEF			
QY	270	AGDTNLRDQEVITKGG	LPDNVFDAWEFLGPKHCQYTWDTRKANNNLRIPAA	YKRRDRIF	329				
Db	269	GGDLNLRDEVS	R---VPDGVKAW	EAAAGSDNKT	KFTWDTF	KNDNKQGFHGA	KMRDRILY	325	
QY	330	FRAEGLHLPQSLDVL	GLEKL-DCGRFPSDHWGLCT	365					
Db	326	W---SGPLDKVKFT	LEGRQIRSCUCF	PSDHWAINAT	359				

RESULT 2
B86252
hypothetical protein [imported] - Arabidopsis thaliana

C:/Species: Arabidopsis thaliana (mouse-ear cress)
C:/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:/Accession: B86252
R:/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:/Reference number: A86141; MUID:21016719; PMID:11130712
A:/Accession: B86252
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-441 <STO>
A:/Cross-references: GB:AE005172; NID:g4835775; PIDN:AAD30241.1; GSPDB:GN00141
C:/Genetics:
A:/Map position: 1

Query Match	14.0%;	Score 275;	DB 2;	Length 441;
Best Local Similarity	31.4%;	Pred. No. 4.2e-15;		
Matches 92;	Conservative 38;	Mismatches 111;	Indels 52;	Gaps

QY	111	ASPSGTPLEDSSTISFITWNIDGDCNLPERARGVCSCLALYSPDVVFLQVPIPYCAY	170
		: : : : : : : :	
Db	163	ASDSTPLTCLKILSYNWFRDLE--LNLRMRAIGHLIQLHSPHLICFQEVTPPEIYDI	219
QY	171	LKKR-----AASYTIITGNEGYFTAILLKKGRVKEFSOEIIPNTKMMRNLLCVNS	224
		: : : : : : : : :	
Db	220	FRKSNMWKAYSCSVSDVAVSRGYCMLLSKLGKVSFSKRS--FGNSIMGRELSIAEVE	276
QY	225	LGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGKMQEAPDSTTVIFAGDTNL	275
		: : : : : : :	
Db	277	VPGRKPLVFATSHLESPCGPPKMQDMFSRERVEQAKAEIAELR--PNANVYFEGDMMW	333
QY	276	RDQEVIKCGS--LPDNVFDAMEFLGPKPKHCQYTWDTKANNNLRIPAYYKHRDRIEFRA	332
		: :	
Db	334	CD----KLDGKRPPLPDKWVDWWEVL-KPGDLGFTYDTKANMMLSGNRALQKRDRILCRL	388
QY	333	EEGHLIPQSLDLVG-----LEKLDGGRPPSDHWGLLCTLN	367
		: : : : :	
Db	389	DDYKL--GGIENVGKEAIPGLSYVEKKVRGDIKKLELPVLPSPDHFGLLVTL	439

RESULT 3
T35506
hypothetical protein SC6E10.19c - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C/Accession: T35506
R/Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A/Reference number: Z21580
A/Accession: T35506
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-253 <SEE>
A/Cross-references: EMBL:AL109661; PIDN:CAB51973.1; GSPDB:GN00070; SCOEDB:SC6E10.19c
A/Experimental source: strain A3(2)
C/Genetics:
A/Gene: SCOEDB:SC6E10.19c

	Query Match	6.0%;	Score 118.5;	DB 2;	Length 253;
	Best Local Similarity	23.7%;	Pred. No. 0.024;		
	Matches 68;	Conservative 39;	Mismatches 115;	Indels 65;	Gaps
QY	107	TILEASP-SGTPLEDSTISFITWNIDGLDGCNLPERARGVCSCLALSPDVVFLOEVIP	165		
		: : : : :		:	
Dd	107	TSLPALPNSTRTEPDGSAVIRVLISYNIRELRD-DTDALARVIKAC---APDIVLLQEAA-P	56		

```

QY      166 PYCAYLK-----RAASTYITITGNEEGYFTAILK-KGRVFKFSQEIIFPNPTKMMRNLL 219
      : : || | : : : || : | : : | : | :
Db      57 RFRFRWRKTTIRLAAAGDLVLLSGGGTAAGPALLCSLRATVERTEDVLLPLTPGRHRRGIA 116
      : : || | : : : || : | : : | : | :
QY      220 CVNVSLGNEECLMTSHLESTREHSAERIRQLKTVLGKMOEAPDSTVYIFAGDTNLRDOE 279
      | : || : : || | : | | : | : | | |
Db      117 AAVVRIGARIGVLSHSLPAD---ERHEQAGLLDHL-AALGVKHAVAGGDLNERP-- 170
      : : || : : || | : | | : | : | | |
QY      280 VIKCG-----LPDNVFDAMEFLGPKHCQYTWDTKANNNLRIPAAYKHHREDRIFFRAEE 334
      || | : : | : | | : || : | : | : |
Db      171 ----GGTRFRRLGEGLRDCW--TAAPWGGEYTFPATAPD-----RRIDAVFVTE-- 213
      : : || : : || | : | | : || : | : |
QY      335 GHLIPQSLDLVGLLEKLDG-----RFPSDHWGLLCTLVN 368
      : : || | : : || | : | | : || : |
Db      214 -----GIEVLGGCVPSGLAGVAEDDLRAATIDHLPVLTALRV 249

```

RESULT 4
D95853
conserved hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) magA
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C/Accession: D95853
R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her
proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A/Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing e
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: D95853
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-252 <KUR>
A/Cross-references: GB:AL591985; PIDN:CAC48492.1; PID:g15139964; GSPDB:GND0167
A/Experimental source: strain 1021, megaplasmid pSymb
R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
C/Genetics:
A/Gene: SMB20092
A/Genome: plasmid

	Query Match	5.7%;	Score 111.5;	DB 2;	Length 252;	
	Best Local Similarity	21.8%;	Pred. No. 0.093;			
	Matches	62;	Conservative	36;	Mismatches 102;	Indels 85; Gaps 11;
QY	124 ISFTWNIDGLDGCNLPERAGVCSCLALYSPDVVFLEOV-----IPPYCAYL	171				
Dd	12 IRELTYNVHSCFGTDRRLDEPARIAAVIAECQPDVIALQEVVDVGARTGGIDAHMIATHL	71				
QY	172 KKRAASYTIITGNEGYFATILLKKGRVKFKSQEIIPENTKMMRNLLCVNVSIGNEFC	231				
Dd	72 NMEAEFHAPALHLEDKEYDAVL--ALPMRLIKAAPLESSEPRGALWEIDAIVAKLQ	128				
QY	232 LMTSHLESTREHSAERIRQLTKTVLGK--MQEAPDSTVIIFAGDTN-----	274				
Dd	129 VIYTHLGL--RGAERLRQAATALLGPWLGMAQGDAHVLAGDLNATGRSTAVRLARQ	185				
QY	275 LRQDEVIKCGGLPNDVEDAMEFLGPKHCQYTWTKANNNULRIPAYKHRFDRIFEFAEE	334				
Dd	186 LSDAQLL-TGVKPRPTFPS-----RLPLL---RIDHV-----	213				
QY	335 GHLLFOSLDLV--GLEKLDC-----GRFPSDHWGLLCTLNVVL	370				
Dd	214 -----LVKGIEVASCRVHGSLARSASDHLPLLAELDDVVM	249				

RESULT 5
H64044

exodeoxyribonuclease III (EC 3.1.11.2) - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Feb-2000
C;Accession: H64044
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kervatage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: H64044
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-267 <TIGR>
A;Cross-references: GB:U32689; GB:LA2023; NID:g1572980; PIDN:AAC21719.1; PID:g1572987; T
C;Superfamily: exodeoxyribonuclease III
C;Keywords: DNA repair; hydrolase

Query Match 5.4%; Score 105.5; DB 2; Length 267;
Best Local Similarity 22.2%; Pred. No. 0.32;
Matches 63; Conservative 43; Mismatches 105; Indels 73; Gaps 14;
QY 124 ISFTWNIDGLDGNLPERARGVCSCLALYSPDVVFLQEV--IPPYCAVLLKKRAASYTII 181
Db 1 MKFISFNINGLRA-----RPHQLAIEIKYQPDVIGLQELKVADEAFPEYTEENLGYHVF 55
QY 182 TGNNEGFTAILLKKGRVKFKSQELIPFPNTKMMRNLCVNSLGNNEFCMLT----- 234
Db 56 HHGQKGHYGVALLTKEPKV-IRRGFTDNEDAQKRIIMADLE--TEFGLLTVINGYFP 111
QY 235 ---SHLESTREHSAERI-RQLKTVLGKMQEAPDSTVIFAGDTNLRDQEV----- 281
Db 112 QGESRAHETKFPRAKEKFYADLQYLER--EHDKSNPILIMGDMNISPSDLIDIGDENRK 169
QY 282 -----KCGGLPD-----NYFDAWEFLGKPKHCQYTW-DTKA-----NNNLRIP 318
Db 170 RMLRTGKCSFLPERAWYQRLYDYGLEDSFRKLNPTANKFSWFDYRSKGFDDNRLRI- 228
QY 319 AAYKHRFDRIFFR--AEEGHLIPQSLDLVGLLEKIDCGRFPSPDH 359
Db 229 -----DHILVSQKLAERCVDVGIALDIRAMEK-----PSDH 259

RESULT 6
F97566
probable nuclease (PA4172) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: F97566
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A;Reference number: A97359; PMID:11743194
A;Accession: F97566
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87487.1; PID:g15156812; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3151
A;Map position: circular chromosome
C;Superfamily: exodeoxyribonuclease III

Query Match 5.2%; Score 102; DB 2; Length 265;
Best Local Similarity 23.5%; Pred. No. 0.63;
Matches 67; Conservative 34; Mismatches 94; Indels 90; Gaps 16;
QY 128 TWNIDGLDGNLPERARGVCSCLALYSPDVVFLQEV--IPPYCAVLLKKRAASYTIIITGNE 185
Db 5 TWMINGVKA-----RIENLCQWLKDDSPDIVCLQELIKSVDEGFPRLIEALGYHIETHGQ 59

QY 186 EGYFTAILLKKGRVKFKSQEI--IPFPNTKMMRNLCVNSLGNNEFCMLTSHLESTRE 242
Db 60 KGFNGVALLS---KVKPDEVNRGLPGDDADEQARFIEGVFSVSGAIVRCSLYLPNGNP 115
QY 243 HS-----AERIRQLKTVLGKMQEAPDSTVIFAGDTNLRDQEVIKCGGLPDNVF 291
Db 116 PDDPVKYPYKLAWMERLRRFAEDRLAMEE-----PLILAGDYNV-----IPE-PF 159
QY 292 D-----AWE----FLGKPK-----HCQYT-----WDTKA-----NNNL 315
Db 160 DCHDPRVWEGDALFLPKTRAFRKLLENLGFTDARATTTDAAGLYSFWDYQAGAWPKNNGI 219
QY 316 RIPAAYKHRFDRIFFRAEEGHLIPQSLDLVGLER-LDCGRFPSPDH 359
Db 220 RI-----DHLMISAEEA-----DRLQSVETIEKHVRAMEKPSDH 252

RESULT 7
AD2787
exonuclease III [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AD2787
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woerage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClair, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AD2787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42714.1; PID:g17740152; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atul714
A;Map position: circular chromosome
C;Superfamily: exodeoxyribonuclease III

Query Match 5.2%; Score 102; DB 2; Length 265;
Best Local Similarity 23.5%; Pred. No. 0.63;
Matches 67; Conservative 34; Mismatches 94; Indels 90; Gaps 16;
QY 128 TWNIDGLDGNLPERARGVCSCLALYSPDVVFLQEV--IPPYCAVLLKKRAASYTIIITGNE 185
Db 5 TWMINGVKA-----RIENLCQWLKDDSPDIVCLQELIKSVDEGFPRLIEALGYHIETHGQ 59
QY 186 EGYFTAILLKKGRVKFKSQEI--IPFPNTKMMRNLCVNSLGNNEFCMLTSHLESTRE 242
Db 60 KGFNGVALLS---KVKPDEVNRGLPGDDADEQARFIEGVFSVSGAIVRCSLYLPNGNP 115
QY 243 HS-----AERIRQLKTVLGKMQEAPDSTVIFAGDTNLRDQEVIKCGGLPDNVF 291
Db 116 PDDPVKYPYKLAWMERLRRFAEDRLAMEE-----PLILAGDYNV-----IPE-PF 159
QY 292 D-----AWE----FLGKPK-----HCQYT-----WDTKA-----NNNL 315
Db 160 DCHDPRVWEGDALFLPKTRAFRKLLENLGFTDARATTTDAAGLYSFWDYQAGAWPKNNGI 219
QY 316 RIPAAYKHRFDRIFFRAEEGHLIPQSLDLVGLER-LDCGRFPSPDH 359
Db 220 RI-----DHLMISAEEA-----DRLQSVETIEKHVRAMEKPSDH 252

RESULT 8
T30299
dynein heavy chain isoform 7A - sea urchin (Tripneustes gratilla) (fragment)
C;Species: Tripneustes gratilla
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30299

[illegible]

A: Molecule type: DNA
 A: Residues: 1-40, 'SGIPEKA', 48-159 <UMA>
 A: Cross-references: GB:U06977
 A: Note: authors translated the codon TCA for residue 41 as Gln, GGA for residue 42 as
 and GCC for residue 47 as Pro
 C: Genetics:
 A: Introns: 29/2; 92/3; 144/1
 C: Superfamily: alpha-2-macroglobulin

```

Query Match      5.0%; Score 98; DB 2; Length 1495;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 66; Conservative 34; Mismatches 106; Indels 82; Gaps

QY      26 AQAEEEDRYKRRRLQCLGFALVGG-----CDPTWPSVLREND--WQTQ-----KALS 70
      ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1149 SQSQESHVYTKALLAYAFALAGNKAKRSELLLESLNKDAVKEEDSLHWQRPDGVQKVALS 1208

QY      71 AYFELPENDQGWPRPPTSPKSEAYV--DLTNEDANDTITILEASPSGTPLEDSSTIS-F 126
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1209 FY-----QPRAPSAEVEMTAYVLLAYLTSSSRPTRLSSS-----DLSTASKI 1252

QY      127 ITW-----NIDGLDGCNLPERARGVCSCLALYSPDVFLQEVIPPYCAVLKKRAASYTII 181
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1253 VKWISKQQNSDG-----GL-----LLPQTLVVALQALSKYGSATFTRSQKEVLV 1296

QY      182 TGNEEGYFTAILLKKGRVKKFSQEI---IP-----FPNTKMMRNLLLCVNS 224
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1297 TSRSSGTFSETFHVNNSGNRLLLOEVRLLPDLPGNYVTKGSGSGCVYLQTSILKYNIL-PVA 1354

QY      225 LGGNEFCMLTSHL-----ESTREHSAERIRQLKTVLGKMQEAPDSTTVI 268
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1355 DGKAPFALQVNTLPLNEDKAEHDHRTFOIRINVSYTG---ERPSSNMVYI 1399

```

```

RESULT 13
B84683
hypothetical protein At2g28300 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84683
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: B84683
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2218 <STO>
A:Cross-references: GB:AE002093; NID:g4803953; PIDN:AAD29825.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28300
A:Map position: 2

```

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 ; Search time 29.9404 seconds
(without alignments)
1646.697 Million cell updates/sec

Title: US-09-697-863A-4
Perfect score: 1968
Sequence: 1 MASGSSSDAAEPAGPAGRAA.....DCGRFPSDHWGLLCTLNVL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1968	100.0	370	21	AAV56020	Mouse CD40 recepto
2	1258	63.9	362	21	AAV56019	Human CD40 recepto
3	1258	63.9	362	22	AAM39841	Human polypeptide
4	1258	63.9	372	22	AAMA1627	Human polypeptide
5	1258	63.9	392	20	AAV03182	Topoisomerase II b
6	1250	63.5	369	21	AAB53403	Human colon cancer
7	1249	63.5	362	22	AAB93674	Human protein sequ
8	1154.5	58.7	306	20	AAV03181	Topoisomerase II b
9	961	48.8	263	22	ABG22067	Novel human diagno
10	647.5	32.9	311	22	ABG22068	Novel human diagno

11	409	20.8	161	20	AAV11868	Human 5' EST secre
12	275	14.0	404	21	AAG26231	Arabidopsis thalia
13	275	14.0	404	21	AAG42551	Arabidopsis thalia
14	275	14.0	426	21	AAG26230	Arabidopsis thalia
15	275	14.0	426	21	AAG42550	Arabidopsis thalia
16	275	14.0	437	21	AAG26229	Arabidopsis thalia
17	275	14.0	441	21	AAG42549	Arabidopsis thalia
18	133.5	6.8	76	20	AAV12175	Human 5' EST secre
19	125.5	6.4	449	22	ABB60974	Drosophila melanog
20	98	5.0	816	22	ABG06343	Novel human diagno
21	95	4.8	474	23	ABP39812	Staphylococcus epi
22	93.5	4.8	551	22	ABG20239	Human inositol pol
23	93.5	4.8	1006	23	AAU98904	Novel human diagno
24	92.5	4.7	319	16	AAAR80147	Chimeric interleuk
25	90	4.6	747	22	ABB71929	Drosophila melanog
26	90	4.6	1011	22	AAB84703	Amino acid sequenc
27	89.5	4.5	750	22	AAU38392	Salmonella typhi c
28	89	4.5	318	18	AAW21746	E3330-binding prot
29	89	4.5	345	21	AAB56610	Human prostate can
30	89	4.5	524	19	AAW48893	Homo sapiens MGMT-
31	89	4.5	945	22	ABG06885	Novel human diagno
32	89	4.5	1284	22	ABG06053	Novel human diagno
33	89	4.5	1284	22	ABG09636	Novel human diagno
34	89	4.5	1284	22	ABG10795	Novel human diagno
35	89	4.5	1284	22	ABG14889	Novel human diagno
36	89	4.5	1703	22	ABG12526	Novel human diagno
37	89	4.5	2253	21	AAV44301	Human acrosome rea
38	88.5	4.5	668	21	AAB28595	Maize Scarecrow pr
39	87.5	4.4	354	22	ABB61565	Drosophila melanog
40	87.5	4.4	577	22	ABG11645	Novel human diagno
41	87	4.4	342	23	ABB48646	Listeria monocytog
42	87	4.4	710	22	ABG20795	Novel human diagno
43	87	4.4	1557	22	ABG02512	Novel human diagno
44	87	4.4	1681	22	ABG01407	Novel human diagno
45	87	4.4	1709	22	ABG02341	Novel human diagno

ALIGNMENTS

RESULT 1	
AAV56020	standard; Protein; 370 AA.
AC	AAV56020;
AC	
DT	15-MAR-2000 (first entry)
XX	
DE	Mouse CD40 receptor associated protein.
XX	
KW	Antiarteriosclerotic; antiarthritic; neuroprotective; dermatological;
KW	immunosuppressive; antinflammatory; immunosuppressive; antiallergic;
KW	human; CD40 receptor associated protein; CRAP; cytoplasmic domain;
KW	tumour necrosis factor; TNF; receptor; superfamily; CD30; homology;
KW	TNF receptor associated factor; TRAF; modulator; signalling pathway;
KW	diagnosis; NF-kappaB; Jun; kinase; atherosclerosis; multiple sclerosis;
KW	arthritis; systemic lupus erythematosus; graft rejection; allergy;
KW	graft versus host disease; autoimmune disease.
XX	
OS	Mus musculus.
XX	
PN	W09955859-A2.
XX	
PD	04-NOV-1999.
XX	
PF	28-APR-1999; 99WO-EP03025.
XX	
PR	29-APR-1998; 98EP-0201392.
XX	
PA	(VIAA-) VIAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX	
PI	Pype SMC, Remacle JEFFG, Huylebroeck DFE;
XX	

DR WPI: 2000-062029/05.
DR N-PSDB; AAZ47119.
XX Novel proteins used to treat inflammatory diseases, NF-kappab related
PT diseases and for improvement of anti-tumor treatments
XX
PS Claim 3; Page 41-43; 48pp; English.
XX
CC This sequence represents the mouse CD40 receptor associated protein
CC (CRAP). CRAP is a functional protein capable of interacting with the
CC cytoplasmic domain of CD40 and/or other receptors of the tumour necrosis
CC factor (TNF) receptor superfamily such as CD30 and TNF receptor I, where
CC the protein has no homology to TNF receptor associated factor (TRAF)
CC proteins. The CD40 binding proteins can be used as modulators of the
CC CD40 signalling pathway, especially to diagnose and treat TRAF-related,
CC CD40-related, NF-kappab related and/or Jun (kinase)-related diseases,
CC and for the improvement of anti-tumour diseases. Diseases which may be
CC treated include atherosclerosis, arthritis, multiple sclerosis, systemic
CC lupus erythematosus, graft rejection, graft versus host disease, allergy,
CC and autoimmune disease. The proteins can be used to sensitize tumour
CC cells to anti-tumour treatments and to screen for compounds which
CC interfere with the interaction of the proteins with other protein
CC components of the TRAF, CD40 or NF-kappab related pathway.
XX
SQ Sequence 370 AA;

Query Match 100.0%; Score 1968; DB 21; length 370;
Best Local Similarity 100.0%; Pred. No. 1.9e-183;
Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGSSSDAEPAGPAGRAASAPPEAAQAEEDRVKRRRLQCLGFALVGGCDPTMPSVLR 60
Db 1 MASGSSSDAEPAGPAGRAASAPPEAAQAEEDRVKRRRLQCLGFALVGGCDPTMPSVLR 60
QY 61 NDMOTOKALSAFYELPENDQGWPRQPTSEKSEAYVDLTNEDANDTILLEASPSGTPLED 120
Db 61 NDMOTOKALSAFYELPENDQGWPRQPTSEKSEAYVDLTNEDANDTILLEASPSGTPLED 120
QY 121 SSTISFITWNIDGLDGCNUPERARVCSCALYSPDVFLQEVIPYCAYLKKRAASYTI 180
Db 121 SSTISFITWNIDGLDGCNUPERARVCSCALYSPDVFLQEVIPYCAYLKKRAASYTI 180
QY 181 ITGNEEGYFTAILLKGRVKFKSQEIIPFPNTKMMRNLLCVNVSLSGNEFCMLTSHLEST 240
Db 181 ITGNEEGYFTAILLKGRVKFKSQEIIPFPNTKMMRNLLCVNVSLSGNEFCMLTSHLEST 240
QY 241 REHSAERIRQLKTVLGKMQEAPDSTTVIFAGDTNLRDQEVIKCGGIPDNVFDWAEFLGKP 300
Db 241 REHSAERIRQLKTVLGKMQEAPDSTTVIFAGDTNLRDQEVIKCGGIPDNVFDWAEFLGKP 300
QY 301 KHQYTWDTKANNNLRIPAAKYHRFDRIFFRAEEGHLIPQSLDLVLEKLDGFRPSDHW 360
Db 301 KHQYTWDTKANNNLRIPAAKYHRFDRIFFRAEEGHLIPQSLDLVLEKLDGFRPSDHW 360
QY 361 GLICTLNVL 370
Db 361 GLICTLNVL 370

RESULT 2
AAY56019
ID AAY56019 standard; Protein; 362 AA.
XX
AC AAY56019;
XX
DT 15-MAR-2000 (first entry)
XX
DE Human CD40 receptor associated protein.
XX
KW Antiarteriosclerotic; antiarthritic; neuroprotective; dermatological;
KW immunosuppressive; antiinflammatory; immunosuppressive; antiallergic;
KW human; CD40 receptor associated protein; CRAP; cytoplasmic domain;
KW tumour necrosis factor; TNF; receptor; superfamily; CD30; homology;

KW TNF receptor associated factor; TRAF; modulator; signalling pathway;
KW diagnosis; NF-kappab; Jun; Kinase; atherosclerosis; multiple sclerosis;
KW arthritis; systemic lupus erythematosus; graft rejection; allergy;
KW graft versus host disease; autoimmune disease.
XX
OS Homo sapiens.
XX
PN WO9955859-A2.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1999; 99WO-EP03025.
XX
PR 29-APR-1998; 98EP-0201392.
XX
PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Pyre SMC, Remacle JEFJG, Huylebroeck DFE;
XX
PT WPI: 2000-062029/05.
DR N-PSDB; AAZ47119.
XX
PT Novel proteins used to treat inflammatory diseases, NF-kappab related
XX diseases and for improvement of anti-tumor treatments

Claim 2; Page 39-41; 48pp; English.

CC This sequence represents the human CD40 receptor associated protein
CC (CRAP). CRAP is a functional protein capable of interacting with the
CC cytoplasmic domain of CD40 and/or other receptors of the tumour necrosis
CC factor (TNF) receptor superfamily such as CD30 and TNF receptor I, where
CC the protein has no homology to TNF receptor associated factor (TRAF)
CC proteins. The CD40 binding proteins can be used as modulators of the
CC CD40 signalling pathway, especially to diagnose and treat TRAF-related,
CC CD40-related, NF-kappab related and/or Jun (kinase)-related diseases,
CC and for the improvement of anti-tumour diseases. Diseases which may be
CC treated include atherosclerosis, arthritis, multiple sclerosis, systemic
CC lupus erythematosus, graft rejection, graft versus host disease, allergy,
CC and autoimmune disease. The proteins can be used to sensitize tumour
CC cells to anti-tumour treatments and to screen for compounds which
CC interfere with the interaction of the proteins with other protein
CC components of the TRAF, CD40 or NF-kappab related pathway.
XX
SQ Sequence 362 AA;

Query Match 63.9%; Score 1258; DB 21; length 362;
Best Local Similarity 68.4%; Pred. No. 4.2e-114;
Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;

QY 24 EAAQAE-EDRVKRRRLQCLGFALVGGCDPTMPSVLRNDMOTOKALSAFYELPENDQGW 82
Db 12 EAAEEGEPEVKRRRLCVFEFASVASCDAVAQCFLAENDWEMERALNSYFEPEESAL 71
QY 83 PROPTSEKSEAYVDLTNEDANDTILLEASPS-GTPLEDSSTISFITWNIDGLDGCNLP 141
Db 72 ERRPETISEPKTYVDLTNEETDSTTSKISPSSEDTQENGSMFSLITWNIDGLDNLSE 131
QY 142 RARGVCSCLALYSPDVFLQEVIPYCAYLKKRAASYTIITGNEGYFTAILLKGRVKF 201
Db 132 RARGVCSYALYSPDVIFLQEVIPYSYLKKRSSNYEITGHEEGYFTAIMLKSRVKL 191
QY 202 KSQEIIPFPNTKMMRNLLCVNVSLSGNEFCMLTSHLESTREHSAERIRQLKTVLGKMQEA 261
Db 192 KSQEIIPFPNTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQLKVLKMQEA 251
QY 262 PDSTTVIFAGDTNLRDQEVIKCGGIPDNVFDWAEFLGKPKHCQYTWDTKANNNLRIPAY 321
Db 252 PESATVIFAGDTNLRDREVTGCGIPNNIVDVWEFLGKPKHCQYTWDTQMSNLGITAC 311

QY 322 KHRFDRIFFR--AEEGHLIPQSLDLVLEKLDGFRPSDHWGLICTLNVL 370
Db 312 KLRFDRIFFRAAEEGHIIPRSLDLGLEKLDGFRPSDHWGLICNLDIIL 362

RESULT 3
AAM39841
ID AAM39841 standard; Protein; 362 AA.
XX
AC AAM39841;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2986.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.
KW
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58997.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2986; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 362 AA;

Query Match 63.9%; Score 1258; DB 22; Length 362;
Best Local Similarity 68.4%; Pred. NO. 4.2e-114;
Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;
QY 24 EAAQAE-EDRVKRRRLQCLGFALVGGCDPTMVPVSLRENDWQTOKALSAVFLEPNDQGW 82

Db 12 FAAEEGEPEVKKRRLLCVFEFASVASCDAVAQCFLAENDWEMERALNSYFEPVEESAL 71
QY 83 PROPPISFKSEAYVDLTNEDANDTTITLEASPS-GTPLEDSSTISFITWNIDGLDGCNLP 141
Db 72 ERREPITSEPKTYVDLTNEETDSTSTKISPSSEDTQENGSMFSLITWNIDGLDNLNSE 131
QY 142 RARGVCSCLALYSPDVVFLQEVIPYCAYLKKRAASYTIITGNEGYFTAILLKKGRVKF 201
Db 132 RARGVCSYLALYSPDVVFLQEVIPYYSYLKKRSSNYEITGHEEGYFTAILMLKSRVKL 191
QY 202 KSGEILPFPNTKMMRNLLCVNVS LGNEFCIMTSHLESTREHSAERIRQLKTVLGKMOEA 261
Db 192 KSGEILPFPSTKMMRNLLCVHVNVS GNELCIMTSHLESTRGHAAERMNQLKVKLKKMOEA 251
QY 262 PDSTTVIFAGDTNLRDQEVTKCGGLPDNVFDWAEFLGKPKHCQYTWDTKANNNLRIPAY 321
Db 252 PESATVIFAGDTNLRDREYTRCGGLPNNIVDVWEFLGKPKHCQYTWDTQMNNSNLGITAAC 311
QY 322 KHRFDRIEFR--AEEGHLIPQSLDLVGLKLDGGRPSDHWGLLCTLANVVL 370
Db 312 KLRFDRIEFRAAAEEGHIIIPRSLDLGLEKLDGGRPSDHWGLLCLNDIIL 362

RESULT 4
AAM41627
ID AAM41627 standard; Protein; 372 AA.
XX
AC AAM41627;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6558.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.
KW
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI60783.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6558; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX WPI, 2000-587534/55.
DR N-PSDB; AAC98160.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11; Page 1508-1509; 2104pp; English.
xx

CC AAC9/9911to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antinfecive and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

SQ Sequence 369 AA;

Query Match	63.5%;	Score 1250;	DB 21;	Length 369;
Best Local Similarity	68.1%;	Pred. No. 2.6e-113;		
Matches 239;	Conservative 43;	Mismatches 65;	Indels 4;	Gaps 3;

24 EAAQAE-EDYKRRRLQCLGIFALVGGCDTMTWPSVLRENDWQTKALSAYFELPENDQGW 82
 19 EAAEEEGEPYKKRRLCVFASVASCAAVAQCLAEINDWEMERALLNSYFEPVVEESAL 78

19 EAAEEGEPEVKKRRLCVFASVASCDAAVAQCFLAENDWEMERALNSYFEPVEESAL 78

[illegible]

QY	142	RARGSYCLALYSPDVFLEQVPIPCAYLKKRASTYITITGNEEGYFTAILLKKGRVYF	201
Db	139	RARGSYLALYSPDVFLEQVPIPCAYLKKRASTYITITGNEEGYFTAILLKKGRVYF	198

QY	202	KSQEIIPEPNTKMRNLLCVANSLGNEFCMLTSHLESTREHSAERIRQLKTVALCKMOEA	261
		: : :	
Dd	199	KSQEIIPEPSTKMRRNLLCVHNVSGNELCLMTSHLESTRGHAAERMOLKMYLKKMOEA	258

QY	262	PDSITVIFAGDTNLRDQEVYKCGGLPDNVFDAMEFLGPKHQYTWDTKANNNLRIPAAAY	321
	1	1	1
	2	2	2
	3	3	3
	4	4	4
	5	5	5
	6	6	6
	7	7	7
	8	8	8
	9	9	9
	10	10	10
	11	11	11
	12	12	12
	13	13	13
	14	14	14
	15	15	15
	16	16	16
	17	17	17
	18	18	18
	19	19	19
	20	20	20
	21	21	21
	22	22	22
	23	23	23
	24	24	24
	25	25	25
	26	26	26
	27	27	27
	28	28	28
	29	29	29
	30	30	30
	31	31	31
	32	32	32
	33	33	33
	34	34	34
	35	35	35
	36	36	36
	37	37	37
	38	38	38
	39	39	39
	40	40	40
	41	41	41
	42	42	42
	43	43	43
	44	44	44
	45	45	45
	46	46	46
	47	47	47
	48	48	48
	49	49	49
	50	50	50
	51	51	51
	52	52	52
	53	53	53
	54	54	54
	55	55	55
	56	56	56
	57	57	57
	58	58	58
	59	59	59
	60	60	60
	61	61	61
	62	62	62
	63	63	63
	64	64	64
	65	65	65
	66	66	66
	67	67	67
	68	68	68
	69	69	69
	70	70	70
	71	71	71
	72	72	72
	73	73	73
	74	74	74
	75	75	75
	76	76	76
	77	77	77
	78	78	78
	79	79	79
	80	80	80
	81	81	81
	82	82	82
	83	83	83
	84	84	84
	85	85	85
	86	86	86
	87	87	87
	88	88	88
	89	89	89
	90	90	90
	91	91	91
	92	92	92
	93	93	93
	94	94	94
	95	95	95
	96	96	96
	97	97	97
	98	98	98
	99	99	99
	100	100	100
	101	101	101
	102	102	102
	103	103	103
	104	104	104
	105	105	105
	106	106	106
	107	107	107
	108	108	108
	109	109	109
	110	110	110
	111	111	111

Dy 322 KHRDRLEFR-AEEGLIPQSLDLVLEKLDGGRPPSDHMGELCTENVVL 370
| | | | | | | | | | | | | | | | | | | | : : : |
Db 319 KLRFRTFFRAAAEGHIIPRLDILGLEKLDCCGRFPSPDHMGELCNDDIIL 369

RESULT 7
AAB93674
ID AAB93674 standard; Protein; 362 AA.

AC AAB93674;

DT 26-JUN-2001 (first entry)

	Human protein sequence	SEQ ID NO:13210.
DE		
VV		

Human; primer; detection; diagnosis; antisense therapy; gene therapy.

05 Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.
PD

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 11-JAN-2000; 2000JP-0118776.
03 MAY 2000 01:03Z

PR 09-JUN-2000; 2000JP-0241899.
XX

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8; SEQ ID 13210; 2537pp + CD ROM; English

CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

50 Sequence 362 AA;

Query Match	63.58;	Score 1249;	DB 22;	Length 362;
Best Local Similarity	68.18;	Pred. No. 3.1e-113;		
Matches 239;	Conservative 43;	Mismatches 65;	Indels 4;	Gaps 3;

QY 24 EAQAE-EDVRKRRRLQGLFALVGGCDPTWPSVLRENDWQTQALSAIFELPENDQGW 82

Db 12 EAAREEGEPEVKKRRLLCYEFASVASCDAAVAQCFLAENDWEMERALNSCFEPVEESAL 71

QY 83 PROPTSFKSEAYVDLTNEDANDTILLEASPS-GTPLEDSSTISFTWNIDGLDGCNLP 141

Db 72 ERPETISEPTYVDLTNEETDSTTSKISPS EDTQÖENGSMFSLITWNIDGLDNLSE 131

QY 142 RARGVCSCLALYSPDVFLQEVIPPYCAYLKKRAASTYITITGNEEGYFTAILLKKGRVK 201
 ||||| |||||:|||||:|||||:| |||||:|||||:||||| |||||
 Db 132 RARGVCSYALYSPDVIFLQEVIPPYCYSLKKRSSNYEITGHEEGYFTAILMLKKSRYKL 191

QY	202	KSQEIPEPNTKMMRNLICVANSIGNEFCIMTSHLESTREHSAERIRQKLTVLGKMOEA	261
		: :: : :	
Db	192	KSQEIPEPSTKMMRNLICVAVNSGNELCIMTSHLESTRGHAERMNQKMLVLKKOEA	251

QY	262	PDSTTVIFAGD	NLRDQEV	IKCGGLPD	NVEDAME	EELGKPK	CQY	TWDTKAN	NNLRIP	AAAY	321
Dd	252	PESATVIFAGD	NLRDREVT	RCGGLP	PNNIVD	WVEFLG	KRKHCQ	YTWDTOM	NSNLGIT	AAC	311

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 275; DB 21; Length 404;
Best Local Similarity 31.4%; Pred. No. 5,1e-18;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

QY 111 ASPSGTPLEDSSTISFITWNIDGLDGCNLPERARGVCSIALYSPDVVFLQEVIPPYCAY 170
DB 126 ASDSGTPLTCLKILSYNWVFRDLE--LNLRMRAIGHLIQLHSPHLICFQEVTPETIYDI 182
QY 171 LKKR-----AASYTIITGNEEGYFTAILKKGRYKFSQETIPFPNTKMRNLLCVNVS 224
DB 183 FRKSNMWMKAYSCSVSDVAVSRGYCMLLSKLGVSFSSKS--FGNSIMGRELIAEVE 239
QY 225 LGGNE-FCIMTSHLES-----TREHSAERIRQKTVLGKMQEAPDSTVI FAGDTNL 275
DB 240 VPGRKPLVFATSHLESPPCPGPKWDOMFSREREVEQAKEAIEILR--PNANVIFGDMNW 296
QY 276 RDQEVIRKCGG--LPDNVFDAMEFLGPKHCQYTWDTKANNNLRIPAAVKHREFRIFA 332
DB 297 CD---KLDGKFPPLPDKWVDWEVL-KPGDLGFTYDTKANPMLSGNRALQKRLDRITCRL 351
QY 333 EEGHLIPQSLDLVG-----LEKIDCGRFPSPDHMGLCTLN 367
DB 352 DDYKL--GCIEMVGKEAIPGISYVKEKKVRGDIKRIELPVLPSDHFGLVTLIS 402

RESULT 13
AAG42551
ID AAG42551 standard; Protein; 404 AA.
XX
AC AAG42551;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 53080.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 275; DB 21; Length 426;
Best Local Similarity 31.4%; Pred. No. 5.5e-18;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

QY 111 ASPSGTPLEDSSTISFTTNIDGLDGCNLPERARGVCSCLALYSPDVVFLOEVIPRYCAY 170
Db 148 ASDSGTPLTCLKILSYNVWFRDLE--LNLRMRAIGHLLIQLHSPHLICRQEVTPETIYDI 204
QY 171 LKKR-----AASYTITGNEEGYFTAILLKGRVKEKSOEIIIPFNTKMMRNLLCVNVS 224
Db 205 FRKSNMWMKAYSCSVSDVAVSRGYYCMLLSKLGVSFSSKS---FGNSIMGRELSTIAEVE 261
QY 225 LGGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGKMQEAPDSTVIYFAGDTNL 275
Db 262 VPGRKPLVFATSHLESPPCPGPKWDQMFSSRREVEQAKAEIILR---PNANVIFGDMNM 318
QY 276 RDQEVYKCGG--LPDNVFDAMEFLGPKPKHCQYTWDTKANNNLRIPAAYKHFDRIFERA 332
Db 319 CD----KIDGKPLPDKWVDWEVL-KPGDLGFTYDTKANPMLSGNRALQKRLDIRLCRL 373
QY 333 EECHLIPQSLDLVG-----LEKIDCGRFPSPDHGGLCTILN 367
Db 374 DDYKL--GGIEMVGKEAIPGLSYVKEKKVRGDIKKLELPVLPSPDHGGLVLTLS 424

RESULT 15
AAG42550
ID AAG42550 standard; Protein; 426 AA.
XX AC AAG42550;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 53079.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 275; DB 21; Length 426;
Best Local Similarity 31.4%; Pred. No. 5.5e-18;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

OY 111 ASPSGTPLEDDSTISFITWNIDGDCNLEPERARGVCSCLALYSPDVFLQEVIPPYCAY 170
Db 148 ASDSGTPLTCLKILISYVWFREDLE--LNLRMRAIGHLIQHLHSPHLICFOEVTPEIYDI 204
OY 171 LKKR-----AASYTTITGNEGYFTAILLKKGRVKFKSQETIIPFNTKMRNLLCVNS 224
Db 205 FRKSNMWKAYSCSVSDVAVSRGYCMLLSKLGKVSFSSKS--FGNSIMGRELSIAEVE 261
OY 225 LGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGKMQEPDSTTVIFAGDTNL 275
Db 262 VPGRRKPLVFATSHLESFPCGPPKWDQMFSSRREVEQAKEAIEILR--PNANVIFGGDMNW 318
OY 276 RDQEVIKCGG--LPDNVFDAMEFLGKPKHCQYTWDTKANNNLRIPAYKHRFDRIFFRA 332
Db 319 CD---KLDGKFLPDKWVDVWEVL-KPGDLGFTYDTKANPMLSGNRAIQKRLDRILCRL 373
OY 333 EEGHLIPQSLDLVG-----LEKIDCGRFPSPDHWGGLCTLN 367
Db 374 DDYKL--GGIEMVGKEAIPGLSYVKEKKVVRGDIKLELPVLPSPDHFGLLVTLIS 424

Search completed: November 14, 2002, 10:27:25
Job time : 30.9404 secs